Cis and trans determinants of epigenetic silencing by polycomb repressive complex 2 in Arabidopsis

Cis- and trans-determinants of epigenetic silencing by Polycomb Repressive Complex 2 in Arabidopsis

Jun Xiao[1], Run Jin[1#], Xiang Yu[1#], Max Shen[1], John Wagner[2], Armaan Pai[2], Claire Song[2], Michael Zhuang[2], Samantha Klasfeld[1], Chongsheng He[3], Alexandre M. Santos[4], Chris Helliwell[5], Jose L Pruneda-Paz[6], Steve A Kay[7], Xiaowei Lin[8], Sujuan Cui[8], Meilin Fernandez Garcia[9], Oliver Clarenz[10], Justin Goodrich[10], Xiaoyu Zhang[4], Ryan S. Austin[11,12], Roberto Bonasio[3] & Doris Wagner[1]*

[1]Department of Biology, University of Pennsylvania, Philadelphia, PA, USA.
[2]Lab course BIOL425 Fall and Spring 2015, Department of Biology, University of Pennsylvania, Philadelphia, PA, USA
[3]Department of Cell and Developmental Biology, University of Pennsylvania, Philadelphia, PA, USA
[4]Department of Plant Biology, University of Georgia, Athens, GA, USA.
[5]CSIRO Agriculture and Food, Canberra, Australia.
[6]Division of Biology, University of California San Diego, La Jolla, CA, USA.
[7] Department of Neurology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA.
[8]College of Life Science in Hebei Normal University, Hebei, P.R. China
[9]Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA, USA.
[10]University of Edinburgh, Edinburgh, UK.
[11]Agriculture & Agri-Foods Canada, London, ON, Canada.
[12]Department of Biology, Western University, London, ON, Canada.

[#]: equal contribution

*Correspondence should be addressed to D.W. (wagnerdo@sas.upenn.edu)

Disruption of gene silencing by Polycomb Complexes leads to homeotic transformations and altered developmental phase identity in plants[1-5]. Here we define short genomic fragments, Polycomb Response Elements (PREs), that direct Polycomb Repressive Complex 2 (PRC2) placement at developmental genes regulated by silencing in Arabidopsis. We identify transcription factor families that bind to these PREs, co-localize with PRC2 on chromatin, physically interact with and recruit PRC2, and are required for Polycomb silencing in vivo. Two of the cis sequence motifs enriched in the PREs are cognate binding sites for the identified transcription factors and are necessary and sufficient for PRE activity. Thus PRC2 recruitment in plants relies in large part on binding of trans-acting factors to cis-localized DNA sequence motifs.

In both the plant and animal kingdoms, Polycomb repression is important for cell identity[2,5-8]. The evolutionarily conserved PRC2 complex trimethylates lysine 27 of histone H3 (H3K27me3), an epigenetic mark that results in compaction of chromatin and silencing of gene expression at thousands of loci[1-5,8,9]. After its establishment, the repressed chromatin state is mitotically heritable[8,10]. Given that PRC2 has no inherent DNA binding specificity, a key question is how the Polycomb epigenetic machinery targets loci it silences. In Drosophila, multiple transcription factors (TFs) bind to *cis* regulatory regions several hundreds of base pairs in length called PREs and recruit Polycomb complexes[8,9]. Despite initial identification of a few such PREs in mammals, recent studies instead implicate promoter proximal unmethylated CpG islands in PRC recruitment[7]. Likewise, PREs with inherent silencing ability were identified at a handful of loci in Arabidopsis[10-14], but it is unclear whether this mechanism broadly underpins PRC2 recruitment.

To elucidate the PRC2 targeting mechanism in Arabidopsis, we identified 132 high-confidence PRC2-regulated genes from our own and public genomic datasets[15-17] (Supplementary Fig. 1a and Supplementary Tables 1 and 2) and computationally defined 170 candidate PREs (600 bp in length) associated with them. We selected five PREs from three loci (the PRC2 targets *AGAMOUS* (*AG*) and *SEPALLATA3* (*SEP3*)[2,18] and a gene of unknown function (*At5g61120* (*At5g*)) to test their ability to recruit PRC2 and direct *de novo* H3K27 trimethylation when randomly integrated into the genome. All five candidate plant PREs recruited PRC2 (represented by complex components FIE, EMF2 and MSI1, Supplementary Fig. 1b) and gained H3K27me3 as did a previously characterized control PRE (PC_LEC2)[12] (Fig. 1 a, b and Supplementary Fig. 2a,b; n=40 transgenic lines). H3K27me3 is known to spread from the site of PRC2 recruitment to adjacent genomic locations[8,9], this was also observed at the PREs (Supplementary Fig. 2a, b). Random unlinked (NC_1 and NC_3) and linked (*AG* locus, NC_2) DNA fragments did not recruit PRC2 and or gain H3K27me3.

In Drosophila, PREs not only recruit Polycomb complexes and become decorated with H3K27me3, they also repress linked genes in a Polycomb dependent manner[8,9]. Likewise, when placed between two constitutive promoters, the five candidate PREs significantly ($P < 0.05$, Mann Whitney *U*-test) silenced three independent reporter genes (GFP fluorescence, beta-glucuronidase activity and herbicide resistance), as did PC_LEC2 (Fig. 1c,d, n=15 and Supplementary Fig. 2c,d, n=15 and Supplementary Fig. 2e, n=60 independent transformants). None of the control DNA fragments had this effect. The ability of the PREs to silence active reporters was dependent on PRC2 in all cases (Fig. 1d, and Supplementary Figs. 2d-e).

To determine which sequence-specific binding proteins associate with the five functional PREs, we performed high-throughput DNA binding assays using a library of 1956 Arabidopsis TFs[19]. Our screen identified 233 PRE-binding TFs (Supplementary Table 3). Fifty-five TFs belonging to 20 families were selected for further characterization on the basis of significant binding to multiple PREs (Supplementary Table 3). Among the 20 PRE-associated TF families, three (sub) families were most significantly enriched (Fig. 2a and Supplementary Table 3): the C2H2 Zinc finger family[20] (C1-2iD ZnF subfamily, 4 of 6 members identified), the plant specific APETALA2-like family[21] (AP2 subfamily, 2 of 6 members identified) and the plant specific BASIC PENTACYSTEINE (BPC) family[22] (class I subfamily, 2 of 3 members identified).

In yeast-two-hybrid tests, more than 50 percent of the 55 PRE-interacting transcription factors - including all of the identified C1-2iD Zn-finger, AP2 and class I BPC subfamily transcription factors - physically interacted with at least one PRC2 complex component (Fig. 2b,c and Supplementary Fig. 3a, Supplementary Table 3). We confirmed contact between select members of the transcription factor subfamilies (TOE1, AZF1and BPC1) and the PRC2 complex (represented by FIE) using bimolecular fluorescence complementation (BiFC) in plant cells (Fig. 2d,e and Supplementary Fig. 3b). No BiFC signal was observed for members of closely related TF subfamilies (Fig. 2d,e). Using plants stably expressing tagged versions of TOE1, AZF1 and BPC1 from their endogenous promoters, we showed that the transcription factors co-immunoprecipitated with the PRC2 complex (FIE) in intact plants (Fig. 2f). Co-immunoprecipitation was also observed in the presence of an endonuclease (Supplementary Fig. 3c).

In a parallel approach, we identified six motifs enriched in the 170 computationally defined PREs with a *de novo* motif analysis pipeline[23] (Fig. 3a,b and Supplementary Table 2). Many of these motifs, in particular the GA repeat and the telobox, had previously been correlated with Polycomb occupancy in Arabidopsis[15,24-27]. To test the biological roles of the GA repeat and telobox motifs in Polycomb silencing, we mutated both motifs in the AG_2 and the At5g PRE (Fig.1). Mutation of GA repeats and telobox motifs significantly reduced the ability of both PREs to silence an active reporter (Fig. 3c). The residual activity of the mutated PREs suggests the presence of additional cis motifs with a role in PRC2 recruitment. The GA repeat and telobox motifs of the AG_2 and At5g PRE are conserved in species of the *Brassicaceae* (Fig.3d). Addition of two GA repeats and teloboxes to a DNA fragment that does not recruit PRC2 (NC_1; Fig. 1b) resulted in a synthetic PRE (NC+), which significantly silenced an active reporter in a PRC2-dependent manner (Fig. 3c). The data suggest that the GA repeat and telobox motifs are necessary and sufficient for PRE activity.

Prior studies had identified GA repeats as the cognate binding sites of class I BPC transcription factors[28] and we confirmed binding of the PRC2 interacting class I BPC TF BPC1 to the GA repeat by electrophoretic mobility shift assay (EMSA) (Supplementary Fig. 4a). We conducted a motif-based DNA interaction screen for TFs that bind the telobox. The screen identified members of the PRE-binding and PRC2-interacting C1-2iD ZnF TF subfamily (Supplementary Fig. 4b,c). Association of the C1-2iD ZnF TF AZF1 with the telobox was verified by EMSA (Fig. 3e). Our findings link sequence motifs important for PRE function to recruitment of TFs that physically interact with PRC2.

We next assessed the genome-wide overlap between chromatin occupancy of PRC2 (FIE) and the two TF families (AZF1 and BPC1) in 30-hour-old plants, a stage when PRC2 function becomes essential for plant development[29]. Significant ($Q<10^{-10}$) binding peaks of FIE, AZF1 and BPC1 co-localized with each other and with H3K27me3 peaks both globally and at individual loci (Fig. 4a,b and Supplementary Fig. 5). 23% of all FIE-bound regions overlapped with BPC1 peaks, while 28% overlapped with AZF1 peaks (Fig. 4c). This overlap was significantly larger than expected by chance for the peak associated genes ($P<10^{-307}$ and $P<10^{-105}$ (hypergeometric test), respectively). In total, 1804 FIE peaks (35%) overlapped with an AZF1 or a BPC1 peak and at 42% of these peaks both TFs were present (Fig. 4d). Cognate binding motifs (GA repeat and telobox) of the class I BPC and the C1-2iD TFs were also significantly enriched under the FIE binding peaks (Supplementary Fig. 5d). Gene Ontology term enrichment analysis links the FIE, AZF1and BPC1 targets to shoot development, flower patterning and gynoecium development (Supplementary Fig. 5e).

Characteristic phenotypes of mutants in the PRC2 methyltransferase CLF are upwards curled leaves with partial floral identity as well as precocious flowering[2]. Higher order mutants in the class I BPC TFs or the C1-2iD C2H2 ZnF TFs do not exhibit these phenotypes[30,31] (Supplementary Fig. 6), suggesting combinatorial roles for the two TF families in Polycomb silencing. Two pieces of evidence support this idea. Firstly, knockdown of either TF family significantly enhanced the leaf curling and the precocious flowering of the hypomorph clf^{R} mutant[18] (Supplementary Figs. 7, 8). Secondly, simultaneous knockdown of both TF families ($BPC{+}ZnF^{KD}$) in the wild type triggered upwards leaf curling and precocious flowering (Fig. 5a,b). The $BPC^{KD}clf^{R}$, $ZnF^{KD}clf^{R}$ and $BPC{+}ZnF^{KD}$ phenotypes were accompanied by a significant reduction in PRC2 (FIE) occupancy and in H3K27 trimethylation at Polycomb target loci and by significant de-repression of the Polycomb targets (Fig. 5c-e and Supplementary Figs. 7-9). FIE occupancy -at peaks with $Q<10^{-10}$ in the wild type- was also reduced genome-wide in 30 hr old $BPC{+}ZnF^{KD}$ plants (Fig. 5f,g and Supplementary Fig. 10). By contrast, occupancy of BPC1 at a Polycomb target locus was not dependent on presence of PRC2 (Supplementary Figs. 11, 12).

Finally, we assessed the contribution of AZF1 and BPC1 to PRC2 recruitment by reciprocal gain of function tests. Tethering both the AZF1 and the BPC1 TF to an artificial promoter in isolated plant cells (protoplasts) triggered levels of FIE recruitment similar to those observed at an endogenous Polycomb target locus (Fig. 6a). In addition, overexpression of BPC1 or AZF1 in *clf*^{R} plants restored PRC2 occupancy and H3K27me3 at Polycomb target loci to near wild-type levels and largely rescued the leaf curling defect of *clf*^{R} (Supplementary Fig. 12). These findings suggest combinatorial roles for class I BPC and C1-2iD ZnF TFs in Polycomb silencing and PRC2 recruitment.

Here we uncover a PRC2 recruitment strategy in Arabidopsis that is strikingly similar to that in Drosophila – including roles for GA repeat-binding and Zn-finger TFs in recruitment (Fig. 6b)[7-9]. Further support for our findings comes from a recent study of PRC2 recruitment to the *ABI4* locus[32]. Our data suggest that a similar logic underpins PRC2 recruitment in species from two kingdoms of life. The plant PREs we uncovered may recruit both PRC2 and PRC1, since GA repeat and telobox motifs also link to PRC1 occupancy in Arabidopsis[26,27]. Additional PREs besides those we predicted likely exist and may act at different stages and in different tissues or conditions. Likewise, additional determinants of PRC2 recruitment remain unidentified. Their discovery combined with the current data should enable computational prediction of PREs for future epigenetic reprograming of cell identity or function to enhance plant growth and yield.

URLs.

R packages used:"vioplot" https://cran.r-project.org/web/packages/vioplot/index.html; leoss smoothing https://stat.ethz.ch/R-manual/R-devel/library/stats/html/loess.html , PCC analysis https://stat.ethz.ch/R-manual/R-devel/library/stats/html/cor.html, PCA analysis https://stat.ethz.ch/R-manual/R-devel/library/stats/html/prcomp.html.

.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGEMENTS

We thank Un-Sa Lee for help with manuscript preparation, students of BIOL425 lab course (Spring 2015 and Fall 2015) for the telobox interactome screen, Jia He for help with fluorometric assay of GUS activity

and Wenhuai Li (Chinese Academy of Sciences) for statistical analyses. Support for the study comes from NSF MCB-1243757 and MCB-1614355 to D.W., NSF IOS-1238142 to X.Z., AAFC GRDI-130 to R.S.A., BBSRC award G20266 to J.G., BBSRC award G20266 and DFG fellowship CL 393/1-1 to O.C., NIH NRSA 1F31GM112417-01 to M.F.G., NIH R01GM056006 to S.A.K. and J.P.P. and R01GM067837, RC2GM092412 to S.A.K., NIH DP2MH107055, MoD 1-FY-15-344, and Searle Scholar Program award to R.B.

AUTHOR CONTRIBUTIONS

D.W. and J.X. conceived of the study and J.X. conducted the majority of the experiments. O.C. and J.G. generated the CLF ChIP-chip dataset. J.P.P. and S.A.K. conducted the high throughput DNA interactome screen. R.A. performed the motif analysis, C.H. contributed to functional PRE analysis and R.J. and M.S. to transcription factor/PRC2 interaction tests. A.P., C.S. and M.Z. identified telobox binding TFs under the guidance of J.W. Cho.H. and R.B. performed library preps. R.B., X.Y., S.K. X.Z. and A.S. conducted bioinformatic analysis. M.F. developed the labeling protocol for EMSA. S.C. and X.L. raised anti-BPC1 antisera. D.W. wrote the paper with input from all authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

References

1. Förderer, A., Zhou, Y. & Turck, F. The age of multiplexity: Recruitment and interactions of Polycomb complexes in plants. in *Current Opinion in Plant Biology* Vol. 29 169-178 (2016).
2. Goodrich, J. *et al*. A Polycomb-group gene regulates homeotic gene expression in Arabidopsis. *Nature* **386**, 44-51 (1997).
3. Mozgova, I. & Hennig, L. The polycomb group protein regulatory network. *Annu Rev Plant Biol* **66**, 269-296 (2015).
4. Pu, L. & Sung, Z.-R. PcG and trxG in plants - friends or foes. *Trends in genetics : TIG* **31**, 252-62 (2015).
5. Xiao, J. & Wagner, D. Polycomb repression in the regulation of growth and development in Arabidopsis. *Current opinion in plant biology* **23**, 15-24 (2015).
6. Lewis, E.-B. A gene complex controlling segmentation in Drosophila. *Nature* **276**, 565-70 (1978).
7. Blackledge, N.-P., Rose, N.-R. & Klose, R.-J. Targeting Polycomb systems to regulate gene expression: modifications to a complex story. *Nat Rev Mol Cell Biol* **16**, 643-9 (2015).
8. Simon, J.-A. & Kingston, R.-E. Occupying Chromatin: Polycomb Mechanisms for Getting to Genomic Targets, Stopping Transcriptional Traffic, and Staying Put. *Molecular Cell* **49**, 808-824 (2013).
9. Kassis, J.-A. & Brown, J.-L. Polycomb group response elements in Drosophila and vertebrates. *Adv Genet* **81**, 83-118 (2013).
10. Sun, B. *et al*. Timing mechanism dependent on cell division is invoked by Polycomb eviction in plant stem cells. *Science* **343**, 1248559 (2014).
11. Lodha, M., Marco, C.-F. & Timmermans, M.-C. The ASYMMETRIC LEAVES complex maintains repression of KNOX homeobox genes via direct recruitment of Polycomb-repressive complex2. *Genes Dev* **27**, 596-601 (2013).
12. Berger, N., Dubreucq, B., Roudier, F., Dubos, C. & Lepiniec, L. Transcriptional regulation of Arabidopsis LEAFY COTYLEDON2 involves RLE, a cis-element that regulates trimethylation of histone H3 at lysine-27. *Plant Cell* **23**, 4065-4078 (2011).
13. Yuan, W. *et al*. A cis cold memory element and a trans epigenome reader mediate Polycomb silencing of FLC by vernalization in Arabidopsis. *Nat Genet* **48**, 1527-1534 (2016).
14. Qüesta, J.-I., Song, J., Geraldo, N., An, H. & Dean, C. Arabidopsis transcriptional repressor VAL1 triggers Polycomb silencing at FLC during vernalization. *Science (New York, N.Y.)* **353**, 485-8 (2016).
15. Deng, W. *et al*. Arabidopsis Polycomb Repressive Complex 2 binding sites contain putative GAGA factor binding motifs within coding regions of genes. *BMC genomics* **14**, 593 (2013).
16. Kim, S.-Y., Lee, J., Eshed-Williams, L., Zilberman, D. & Sung, Z.-R. EMF1 and PRC2 cooperate to repress key regulators of Arabidopsis development. *PLoS Genet* **8**, e1002512 (2012).
17. Zhang, X. *et al*. Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. *PLoS Biology* **5**, 1026-1035 (2007).
18. Lopez-Vernaza, M. *et al*. Antagonistic roles of SEPALLATA3, FT and FLC genes as targets of the polycomb group gene CURLY LEAF. *PLoS One* **7**, e30715 (2012).
19. Pruneda-Paz, J.-L. *et al*. A genome-scale resource for the functional characterization of Arabidopsis transcription factors. *Cell Rep* **8**, 622-632 (2014).
20. Kodaira, K.-S. *et al*. Arabidopsis Cys2/His2 zinc-finger proteins AZF1 and AZF2 negatively regulate abscisic acid-repressive and auxin-inducible genes under abiotic stress conditions. *Plant Physiol* **157**, 742-756 (2011).
21. Kim, S., Soltis, P.-S., Wall, K. & Soltis, D.-E. Phylogeny and domain evolution in the APETALA2-like gene family. *Mol Biol Evol* **23**, 107-120 (2006).
22. Monfared, M.-M. *et al*. Overlapping and antagonistic activities of BASIC PENTACYSTEINE genes affect a range of developmental processes in Arabidopsis. *Plant J* **66**, 1020-1031 (2011).

23. Winter, C.-M. *et al*. LEAFY target genes reveal floral regulatory logic, cis motifs, and a link to biotic stimulus response. *Developmental cell* **20**, 430-43 (2011).
24. Molitor, A.-M. *et al*. The Arabidopsis hnRNP-Q Protein LIF2 and the PRC1 subunit LHP1 function in concert to regulate the transcription of stress-responsive genes. *The Plant Cell*, tpc.00244.2016 (2016).
25. Wang, H. *et al*. Arabidopsis Flower and Embryo Developmental Genes are Repressed in Seedlings by Different Combinations of Polycomb Group Proteins in Association with Distinct Sets of Cis-regulatory Elements. *PLoS Genetics* **12**, 1-25 (2016).
26. Zhou, Y., Hartwig, B., Velikkakam James, G., Schneeberger, K. & Turck, F. Complementary activities of TELOMERE REPEAT BINDING proteins and Polycomb Group complexes in transcriptional regulation of target genes. *The Plant Cell* **28**, TPC2015-00787-RA (2015).
27. Hecker, A. *et al*. The Arabidopsis GAGA-Binding Factor BASIC PENTACYSTEINE6 Recruits the POLYCOMB-REPRESSIVE COMPLEX1 Component LIKE HETEROCHROMATIN PROTEIN1 to GAGA DNA Motifs. *Plant physiology* **168**, 1013-24 (2015).
28. Kooiker, M. *et al*. BASIC PENTACYSTEINE1, a GA binding protein that induces conformational changes in the regulatory region of the homeotic Arabidopsis gene SEEDSTICK. *Plant Cell* **17**, 722-729 (2005).
29. Bouyer, D. *et al*. Polycomb repressive complex 2 controls the embryo-to-seedling phase transition. *PLoS Genetics* **7**(2011).
30. Ciftci-Yilmaz, S. & Mittler, R. The zinc finger network of plants. *Cell Mol Life Sci* **65**, 1150-1160 (2008).
31. Meister, R.-J. *et al*. Definition and interactions of a positive regulatory element of the Arabidopsis INNER NO OUTER promoter. *Plant J* **37**, 426-438 (2004).
32. Mu, Y. *et al*. BASIC PENTACYSTEINE Proteins Repress ABSCISIC ACID INSENSITIVE4 Expression via Direct Recruitment of the Polycomb-Repressive Complex 2 in Arabidopsis Root Development. *Plant Cell Physiol* **58**, 607-621 (2017).

Figure legends

Figure 1 Identification of *Arabidopsis* DNA fragments with PRE activity.
(**a**) Construct to test ability of candidate PREs or control DNA fragments (NC) to recruit PRC2 and H3K27me3. Below: region tested by ChIP-qPCR.
(**b**) Top: Occupancy of the PRC2 component FIE at fragments tested. Bottom: H3K27me3 accumulation relative to H3 at fragments tested. Shown are mean ± SEM of three independent ChIP experiments (red dots). **, P< 0.01; ns, not significant P>0.05 one-tailed unpaired *t* test relative to NC_1. The *LEC2* PRE[12], serves as positive control (PC_LEC2). See also Supplementary Fig. 2a,b.
(**c**) Construct to test ability of PREs to silence active reporters. Candidate PREs and control fragments were placed between two constitutive promoters (pF3H and pMAS) driving expression of GFP, GUS or an herbicide resistance gene (Bar), respectively.
(**d**) GFP intensity visually scored (1 = no reporter expression to 5 = full reporter expression) in 15 independent transformants in the wild type (WT) (top) or a *prc2* mutant (*clf-28 swn-7*) (bottom). Violin plot of GFP intensities of presumptive PREs (left) and negative controls (right): range, median= white circle, mean= white line, lower to higher quartile= vertical black line. Black bar: median GFP fluorescence of the PRE populations in the wild type background. *, P<0.05; **, P<0.01; ns, not significant (P>0.16) relative to NC_1, one-tailed Mann-Whitney *U* test. Similar results were obtained for silencing of beta-glucuronidase activity and herbicide resistance (Supplementary Fig. 2c-e).

Figure 2 PRE-binding TFs physically interact with PRC2.
(**a**) TF families most enriched as PRE-binding based on high-throughput yeast-one-hybrid assays. Y-axis: P-value hypergeometric test. Inset: Most enriched TF subfamilies.
(**b, c**) Interaction between PRE-binding TFs and PRC2 by medium throughput yeast-two-hybrid test for all 55 PRE-binding TFs (b) and quantitative ortho-Nitrophenyl-β-galactoside (*ONPG*) assays for members of the most enriched TF subfamilies (c). Negative controls: LFY TF and AD alone. Mean ± SEM of three independent yeast-two-hybrid experiments (red dots). *, P<0.05; **, P< 0.01; ns, not significant P> 0.05 relative to BD, one-tailed unpaired *t*-test.
(**d, e**) Interaction between PRE-binding TFs and the PRC2 complex (represented by the single copy PRC2 component FIE) by bimolecular fluorescence complementation in protoplasts. TFs tested (subfamily): TOE1 (AP2), AZF1 (C1-2iD Zn Finger) and BPC1 (class I BPC) and controls: BBM (AP2/ANT), ZAT5 (C1-2iC Zn Finger), BPC6 (class II BPC) plus LFY. Nuclear fluorescence (d) and quantification (e). Bar =10 µm. Box and whisker plot with median of three BiFC experiments comprising 150 cells each (red line), upper and lower quartile (box) and minima, maxima (whiskers). P-value: one tailed Mann-Whitney *U* test, ***, P<0.001 relative to controls.
(**f**) Co-immunoprecipitation of HA-tagged PRC2 (FIE) after IP of MYC tagged TFs in seedlings. TRB2: negative control. % input values represent mean ± SEM of relative IP (normalized to input FIE-HA and TF levels) from three experiments. Supplementary Fig. 3 shows co-IP in the presence of an endonuclease.

Figure 3 Cis motifs enriched in plant PREs required for PRE activity.
(**a, b**) Sequence logos for position-specific scoring matrices (PSSMs) of motifs identified by *de novo* motif analysis (a) and their enrichment (p-values; converted from the Z-scores of the motif enrichment calculations using a normal distribution) and frequency (%) in the 170 candidate PREs (b).
(**c**) Test of GA repeat and telobox function. Top: PRE Diagram indicating GA repeats and telobox motifs mutated in PREs or added to NC_1. Below: Effect of mutation of GA repeat and telobox motifs in the AG_2 and At5g PREs (AG_2_mu or At5g_mu) or insertion of GA repeat and telobox motifs into the negative control fragment NC_1 (NC+) on PRE activity in populations of independent primary transformants (n=15) in the wild type (WT) or the *prc2* mutant (*clf-28*). Violin plot: range, median= white circle, mean= white line, lower to higher quartile= vertical black line. *, p<0.05; **, p<0.01; one-tailed Mann-Whitney *U* test.

(**d**) Evolutionarily conservation of GA repeats and telobox motifs in the AG_2 PRE and At5g_PRE in *Brassicaceae* species. Vertical lines: additional PRE sequences omitted, *fully conserved nucleotide; : and . partially conserved nucleotide.
(**e**) Electrophoretic mobility shift assay (EMSA) to test association of the AZF1 C1-2iD ZnF TF with the telobox but not mutant versions thereof. % complex: fraction of shifted DNA (mean ± SEM) from three independent EMSA experiments. HIS-TF, unrelated TF control.

Figure 4 ChIPseq analysis to test chromatin occupancy of FIE, H3K27me3, AZF1 and BPC1 in 30-hour-old plants.
(**a**) Heatmap of background-corrected ChIP enrichment for PRC2 (FIE), H3K27, AZF1 and BPC1. Significant peaks ($Q<10^{-10}$) were centered on FIE peak maxima and rank ordered from highest (top) to lowest (bottom) FIE binding peak significance. For each factor and mark, three independent ChIPseq experiments were performed and sequenced as were three matched input controls. The highly consistent ChIP replicates were normalized by sequencing depth and averaged.
(**b**) Browser view of input subtracted ChIPseq signals at *AG*, *LEC2* (PC) and *ACT2* (NC_3) and additional FIE-bound loci. Significant peaks ($Q < 10^{-10}$) according to MACS2 are marked by horizontal bars, with the black saturation proportional to the Q value (as for the narrowPeak file format by ENCODE). As previously reported for seedlings[15] about half of all FIE peaks overlapped with H3K27me3.
(**c**) Percent ChIPseq overlap (by row) for the significant ($Q < 10^{-10}$) FIE, AZF1 or BPC1 peaks with other significant peaks. Shading indicates strength of overlap.
(**d**) Fraction of significant ($Q < 10^{-10}$) FIE peaks that overlap with other significant peaks.

Figure 5 Class I BPC and C1-2iD ZnF TF families are required for Polycomb silencing and PRC2 recruitment *in planta*.
(**a, b**) Flowering time and leaf curling in double BPC and ZnF family knockdown plants ($BPC{+}ZnF^{KD}$) compared to wild type (WT), hypomorph clf^{R} and null *clf-50* mutants. Bar = 1cm. (b) Quantification of leaf curling. Box and whisker plot with median (red line; n=15 plants), upper and lower quartile (box) and minima, maxima (whiskers). Different letters above bars indicate significantly different groups, $p<0.05$ based on Kruskal-Wallis test with Dunn's posthoc test.
(**c-e**) PRC2 (FIE) occupancy, H3K27me3/H3 accumulation and gene expression in genotypes described in (a) and in BPC^{KD} clf^{R} or ZnF^{KD} clf^{R} plants. Negative control loci: *ACT2* for ChIP, *EIF4* for gene expression. Expression in the mutant lines is shown relative to the wild type. Mean ± SEM from three independent experiments (red dots). Black asterisks- significantly different from WT; grey asterisks – significantly different from clf^{R}. * $P<0.05$; ** $P< 0.01$, ns $P>0.05$; one-tailed unpaired student *t* test.
(**f, g**) ChIPseq analysis of FIE occupancy in wild type (WT) and $BPC{+}ZnF^{KD}$ plants. For each factor and mark, three independent ChIP experiments were sequenced as were three matched input controls. The highly consistent ChIP replicates were normalized by sequencing depth and averaged. (f) FIE occupancy ($Q < 10^{-10}$) region metaplot in wild type (WT) and $BPC{+}ZnF^{KD}$ samples, (g) screenshots of input subtracted FIE occupancy. red: previous FIE ChIPseq (Fig. 4), blue: FIE ChIPseq data in WT (center) and in $BPC{+}ZnF^{KD}$ (bottom).

Figure 6 Tethering class I BPC and C1-2iD ZnF TF family to the DNA and test of PRC2/FIE recruitment.
(**a**) Tethering of BPC1 (BPC-LexA DBD) and AZF1 (ZnF-Gal4 DBD) to an artificial promoter (LG_GUS) and test of FIE recruitment using anti –FIE antibody for ChIP qPCR in protoplasts. Endogenous loci tested: *AGAMOUS* (AG_2, positive control (PC)) and *ACT2* (negative control (NC)). Empty vector (EV) and VP16 serve as control tethering vectors. Shown are mean ± SEM of three ChIP experiments (red dots). * $P<0.05$; ** $P< 0.01$; relative to empty vector (EV), one-tailed unpaired *t* test. (**b**) Model for PRC2 recruitment. See text for details.

METHODS:

Plant material and treatment

Mutants and transgenic plants previously described: *clf-28*, *clf-28 swn-7*[33]; *clf-50*, *clf*^{R} (pCLF:CLF-GR *clf-50*)[18]; *bpc123*[22] pFIE:FIE-HA *fie-11*[34]; pEMF2:EMF2-3XFLAG *emf2*[35]; pMSI1::GFP-MSI1 *msi1-1*[36]; 35S::GFP-CLF *clf-50*[37]. In *clf*^{R}, the *clf-50* RNA null mutant is partly rescued by 'leaky' nuclear translocation of pCLF:CLF-GR in the absence of steroid treatment. *clf*-50 and *clf*^{R} are in the Ws accession, pFIE:FIE-HA *fie-11* is in the C24 accession, all other plants are in the Col-0 accession.

PRE tests

To test PRC2 recruitment and H3K27me3 by PREs or control fragments, progeny pools of 40 random T1 plants were analyzed. This strategy was adopted to minimize outliers caused –for example- by position effect. Independent pools tested gave similar results. Control fragments included a known PRE (PC_LEC2) and three random DNA fragments (NC_1, intron of At1g60200; NC_2, promoter of AG; NC_3, 3'UTR of Actin 2 (At3g18780)). GFP intensity was scored visually using a dissecting fluorescence microscope (Olympus, MVX10). A fluorometric (MUG) assay was used to quantify beta-glucuronidase (GUS) activity as previously described[38,39]except that the 4-MU produced was normalized over the fresh weight of each plant. GFP and GUS reporter silencing was assayed in independent primary transformants (T1 plants). For the herbicide resistance assay, primary transformants were transplanted into soil after selection and sprayed with the Basta herbicide (200mg/l) (Bayer Crop Science) 2-3 days after transplanting. Survival rate was scored 5-7 days later.

Transgenic plants

Candidate PRE DNA fragments (~600bp in length) were cloned into pFK205[40] and transformed into wild type (Col-0), pMSI1:GFP-MSI1 *msi1-1* or pEMF2:EMF2-3XFLAG *emf2* for chromatin immunoprecipitation (ChIP) assays. Plants were selected on 1/2 Murashige and Skoog (MS) medium[41] (Sigma) with 40mg/l kanamycin.

A dual reporter system (pPRE-dual-rep) was generated containing the FLAVANONE 3-HYDROXYLASE regulatory region plus the 35S minimal promoter (pF3H-35S mini) driving expression of GFP and beta-glucuronidase and the mannopine synthase promoter (pMAS) driving the BAR gene. Candidate PRE or control fragments were cloned into pPRE-dual-rep and transformed into wild type (Col-0) or *prc2* mutants (*clf*-28, *clf-28 swn-7*/+). PRE fragments for test of loss- or gain- of GA repeat and telobox motifs were synthesized (see Supplementary Table 5) (GenScript Inc. Company, Piscataway,

NJ, US) and shuffled into pPRE-dual-rep. Plants were selected on 1/2 MS plates with 25mg/l hygromycin.

For TF knockdown, 300 bp regions conserved in the PRC2 recruiting TF subfamilies (e.g. C1-2iD Zn finger) but not in the larger TF family (C1-2i Zn finger) were PCR amplified and inserted into vector pRNAi-GG. Plasmids were introduced into *A. tumefaciens* strain GV3101 and transformed into clf^{R} and wild-type plants using floral dip[42]. T1 plants were selected on 1/2 MS plates with 40mg/l kanamycin. BPC^{KD} plants were crossed with ZnF^{KD} plants to generate double knockdown lines.

For overexpression of BPC1 or AZF1 in clf^{R}, the cDNA of BPC1 and AZF1 was cloned into vector pGWB12[43]. Plasmids were introduced into *A. tumefaciens* strain GV3101 and transformed into clf^{R} using floral dip[42]. T1 plants were selected on 1/2 MS plates with 40mg/l kanamycin and 20mg/l hygromycin.

For TF CoIP and ChIP, genomic fragments spanning the upstream intergenic region and the coding region for each TF were PCR amplified [BPC1 (-3225 to +849 bp), AZF1 (-1628 to +735 bp), TOE1 (-8685 to +1353 bp), TRB2 (-1194 to +1227 bp)], cloned into pEG303[44] and transformed into pFIE:FIE-HA *fie-11* for CoIP and into wild type (Col-0) for ChIP.

Phenotype quantification and qRT-PCR

Plants were grown at 22□ in short day conditions (8 hr light/ 16hr dark, light intensity: ~140 μMol/m^2s^1). The length and width of the blade of the fifth rosette leaf was measured for > 15 plants at day 20. RNA was extracted from 3- to 7-day-old plants grown in short day condition and quantitative real-time PCR was performed as previously described[45].

Identification of CLF-binding sites using ChIP-chip

ChIP was conducted with anti-GFP antibodies (Molecular Probes) in 35S::GFP-CLF *clf-50* plants[37] as previously described[17]. Four independent replicates were performed. Amplification, labeling and microarray hybridization was as reported[46]. Genomic regions enriched for CLF binding sites were identified by comparing CLF ChIP-chip to input DNA (4 replicates) using the Tilemap program with the Hidden Markov model (HMM) option[47]. Adjacent probes with HMM posterior probabilities of p>0.5 or higher were merged into regions by requiring a minimal run of 50 bp and allowing a maximal gap of 200 bp, and TAIR5 coordinates were converted to TAIR10 coordinates using the "update_coordinates.pl" script from TAIR. The 3,648 regions enriched for CLF binding are listed in Supplementary Table 1.

Identification of candidate PREs

Putative PRE-containing regions were identified based on the following two conservative criteria. First, we compared previously published genome-wide distribution of H3K27me3, EMF1 and FIE[15-17] with that of CLF generated in this study. EMF1 is a putative PRC1 component that frequently co-localizes with PRC2[16]. A candidate PRE-containing region was required to overlap with at least 3 of the 4 datasets (to account for potential false negative results in these dataset and the redundant contribution of other PRC components). A total of 1,504 regions were identified and assigned to 851 Arabidopsis genes based on previously described criteria[23]. Second, we required target genes associated with these putative PREs be expressed in a highly tissue-specific manner or be de-repressed in *prc2* mutants (*clf swn* and *fie*)[15,33]. For the former, we analyzed previously published transcription profiles in different tissues[48], and defined tissue-specifically expressed genes as those with expression levels higher than 5x the baseline levels. Since all ChIP data (FIE, CLF, EMF1, H3K27me3) was from vegetative development, baseline was defined as the mean expression in samples ATGE_7; ATGE_87; ATGE_12; ATGE_26; ATGE_1; ATGE_19; ATGE_15; ATGE_13; ATGE_20; ATGE_21; ATGE_14; ATGE_17; ATGE_18; ATGE_91; ATGE_5; ATGE_16; ATGE_11; ATGE_10 from AtGen Express[48]. PRE linked genes de-repressed in *clf swn* or *fie* were defined as in[15,33]. The combined expression filters resulted in identification of 132 high-confidence PRC2-regulated genes. 170 candidate PREs were associated with the 132 PRC2-regulated genes. Five of the candidate PREs were selected for *in planta* PRE tests.

Motif prediction and mapping

De novo motif prediction was performed as previously published with minor changes[23,49-51]. Briefly, we applied the motif prediction pipeline to a subset of the candidate PREs (those bound by CLF, EMF2 and marked by H3K27me3, 70 PREs) to enhance prediction performance and lower false positive rates[52]. Motifs were subsequently tested for enrichment within the entire PRE set. Motif width was set from 6-16bp when applicable. The background set consisted of 600bp genic terminal regions from the TAIR10 genome[15]. Motif enrichment was calculated using a non-parametric deterministic sampling as described previously ($Z \geq 3$; $p < 0.027$)[23]. Highly degenerate motifs were filtered from the results using an information quality statistic ($IQ > 20$) defined as:

$$IQ_L = \sum_L N_{L,A} log_2(\frac{F_{L,A}}{P_A})$$

where $N_{L,A}$ is the count of residue type A at position L, $F_{L,A}$ is the frequency of occurrence of the residue of type A at position L in the PSSM, and P_A is the background frequency expected for residue A.

Predicted PSSMs were aligned by position of maximum average site-wise Euclidean distance and hierarchically clustered in the R statistical programming environment. A representative consensus PSSM was then chosen from the PSSMs in each clade, along with a merged candidate, by maximizing for significance score and frequency within target genes. A functional-depth cutoff was used in mapping PSSMs to PREs and to genomic and ChIP data sequence sets (GAGA: 0.62, CTCC: 0.42 and CCG: 0.8 and 0 for telobox, CAA repeats and G-box). When mapping GAGA motif occurrences, overlapping and nearby matches were merged into a common region using the BEDTools merge function with a maximum distance between features of 8bp[53].

Yeast-two-hybrid tests

EMF2, CLF, MSI1, and FIE were cloned into pDEST32 (Clontech) and introduced into yeast strain AH109 (MAT a). Full-length clones were used except for EMF2 and CLF. EMF2C is a better interactor in yeast, while CLFN overcomes the growth defects caused by full-length CLF[54]. The 55 PRE-interacting TFs were cloned into pDEST22 (Clontech) and transfected into yeast strain Y187 (MAT α). Protein−protein interactions were tested after mating as described in the Matchmaker protocol (Clontech). Interaction strength was quantified for a subset of TFs using ortho-nitrophenyl-β-D-galactopyranoside (ONPG) assays[55] for nine independent colonies for each interaction pair in three pools. In addition, different TF fragments (AZF1-N: 1-90 aa; AZF1-M: 91-194 aa; AZF1-C: 195-245 aa; BPC1-N: 1-140 aa; BPC1-C: 141-283 aa; TOE1-N: 1-151 aa; TOE1-M: 152-310 aa; TOE1-C: 311-464 aa) were cloned into pDEST22 and co-transformed with EMF2C or CLFN into yeast strain AH109 followed by scoring of yeast growth.

Bimolecular Fluorescence Complementation (BiFC) assays

FIE and all TFs were cloned into pUC-SPV-NE^{GW} and pUC-SPV-CE^{GW} constructed by shuffling the split Venus-Gateway cassette from pDEST-VYNE/CE(R)^{GW} vectors[56] into pUC18. BiFC assays in Arabidopsis protoplasts were conducted and visualized as previously described[57,58]. For each experiment, fluorescence was compared in protoplast populations prepared and transfected at the same time. Three independent BiFC experiments were performed for each combination of factors tested with at least150 protoplasts scored per replicate. Representative images were taken with a confocal microscope with the same gain (Leica, LCS SL).

CoIP

Co-IP was performed as described[59] with some modifications. Myc-gBPC1, Myc-gAZF1, Myc-gTOE1 or Myc-gTRB2 (TELOMER REPEAT BINDING 2 protein, negative control) were transformed into

pFIE:FIE-HA *fie-11* plants. Three-day-old double transgenic seedlings were harvested after growth in long day conditions. Tissue was ground in protein extraction buffer (20 mM Tris-HCl pH8.0, 150 mM NaCl, 1 mM EDTA, 10% Glycerol, 0.2% Triton X-100, 1 mM PMSF, 1×Protease inhibitor (Roche)), filtered, and centrifuged. The supernatant was incubated with anti-Myc antibody (C3965 or 05-724, Sigma) coupled to protein A Dynabeads (ThermoFisher) overnight at 4 degrees. Beads were washed four times with wash buffer, and bound proteins eluted with elution buffer containing 2% SDS for immunoblotting with HA-HRP conjugated antibody (3F10, Roche). In some reactions Benzonase endonuclease (E1014, Sigma) was added into the protein sample (20 U) and incubated on ice for 1hr prior to immunoprecipitation. Full length gen images for Western analyses after co-IP and for protein abundance in BiFC experiments are sown in Supplementary Figure S13.

ChIP

ChIP was performed as previously described[60], with minor modifications. Extraction buffer I (0.4M Sucrose, 10mM Tris-HCl pH8.0, 10mM MgCl2, 5mM b-ME, 1mM PMSF, 1X Protease inhibitor (Roche)) and II (0.25M Sucrose, 10mM Tris-HCl pH8.0, 10mM MgCl2, 1% Triton X-100, 5mM b-ME, 1mM PMSF, 1X Protease inhibitor (Roche)) were used for protein extraction. The following antibodies that had previously been used for ChIP in Arabidopsis were employed: anti-GFP (A6455, Thermo Fisher)[61] for pMSI1:GFP-MSI1 *msi1-1* ChIP, anti-HA (12CA5, Roche)[61] for pFIE:FIE-HA *fie-11* ChIP, anti- H3K27me3 (07-449, Millipore)[33,62], anti-H3 (07-690, Millipore)[62], anti-FLAG (F3165, Sigma)[63] for pEMF2:EMF2-3XFLAG *emf2* ChIP and anti-Myc (C3956, Sigma)[64] for gAZF1-Myc and gBPC1-Myc ChIP. Anti-FIE[34] antiserum was first used for ChIP here and showed much reduced occupancy in *prc2* mutants (Fig. 5c). The anti-BPC1 antiserum was generated in rabbits using full length recombinant BPC1 protein and gave ChIP signal specifically in the wild type (Supplementary Fig. 11). Throughout H3K27me3 was normalized over H3 to control for nucleosome density, all other ChIP reactions are shown as % input. For most ChIP experiments, representative transgenic lines were used, alternatively pools of 40 independent T1 progeny (or more) were used.

ChIP-seq in germinating embryos

For H3K27me3, FIE, AZF1 and BPC1 ChIPseq, ChIP was performed on germinating embryos (30 hrs after imbibition) as described above, but eluted into a smaller volume (15ul in total), using the Qiagen MinElute PCR purification kit (Cat. No. 28004). Three independent ChIP and input reactions were sequenced. ChIP-DNA and input (after dilution to 0.1~1 ng) were amplified using the SeqPlex DNA Amplification kit (Sigma, SEQXE-10RXN) according to the manufacturer's instructions with the following modifications: a first linear PCR was followed by a second round of amplification (<10 cycles).

After primer removal, qPCR was performed to test amplification of different genomic regions. Linearly amplified DNA from input chromatin and pull-downs was converted to libraries for sequencing by performing end-repair followed by A-tailing and ligation of universal adapters (all enzymes by Enzymatics, MA). Libraries were amplified to 50–100 nM using custom dual indexing primers and sequenced with an Illumina NextSeq500 at a depth of > 15 million reads per sample for pull-downs and > 30 million reads per sample for inputs. We mapped reads to the TAIR10 genome release v31 using bowtie2[65]. Alignment files were converted to 1-bp resolution bigwig files and normalized by 10 million reads sequenced (RP10M) using custom scripts. All ChIPseq replicates were highly similar to each other (see below). Bigwigs from the ChIPseq replicates were averaged using WiggleTools[66]. Significant peaks were identified with MACS2 version 2.1.1.20160309 using relevant input controls (C24 for FIE, Col for AZF1 and BPC1, total H3 pull-down for H3K27me3). Default MACS2 settings were used for FIE, AZF1, and BPC1, and the "--broad" option was used for H3K27me3. Only peaks with Q-value < 10^{-10} were considered.

Significant peaks were mapped to genes as previously described[23] using the HOMER[67] script annotatePeaks. Gene Ontology (GO) enrichment analysis was performed using AgriGO[68] combined with manual curation to remove redundant terms.

For FIE ChIPseq in WT and double TF subfamily (class I BPC and C1-2iD ZnF) knock-down lines, anti-FIE[34] antiserum was used. ChIP-DNA and input libraries (three replicates each) were generated using the ThruPLEX DNA-seq kit (RUBICON GENOMICS, cat. R400406). Libraries were sequenced at a depth of > 15 million reads per sample for pull-downs and > 30 million reads per sample for inputs. Replicate comparison and significant peak identification was as described above using default MACS2 settings. Only peaks with Q < 10^{-10} were considered.

To assess the change in FIE binding in *BPC + ZnF* ^{KD} mutants relative to the wild type (WT), normalized reads mapping to WT FIE peaks (Fig. 5f) were extracted from WT and *BPC + ZnF* ^{KD} FIE ChIP datasets using bedtools and plotted [$\log_2$(RPM)] in a scatterplot. Reads increased or decreased in *BPC + ZnF* ^{KD} were indicated (P-value <0.01 and 2-fold change). For the region metaplot, detected FIE peaks were quantified using the normalized and input-subtracted number of reads in each 10 bp window (FIE peak signal) in WT and *BPC + ZnF* ^{KD}. The average FIE peak signals in the window of +/- 1000 bp around peak centers were plotted with the "loess" smoothing function in R with span = 0.2.

ChIP replicate concordance was assessed by computing pairwise Pearson Correlation Coefficients (PCC) of the RP10M normalized read counts in each 10 kb window for all ChIPseq datasets using the "cor" function with method "pearson" in R. PCC was > 0.95 for all ChIPseq replicates. Principal component analysis (PCA) was performed on the pairwise PCC matrix using the "prcomp" function in R and converted to a scatterplot. Significant regions for all ChIPseq experiments performed in 30-hr-old plants (germinating embryos) are listed in Supplementary Table 4.

EMSA

A 60 bp fragment of the AG_2 PRE containing two GA repeats or one telobox, as well as versions thereof with motif substitutions were labeled with Cy5-dCTP (GE Healthcare Life Sciences) by end-repair with Klenow Fragment (3′→ 5′ exo-) (NEB). For EMSA and other oligonucleotide sequences see Supplementary Table 5. Briefly, complementary single strand DNA probes were synthesized (IDT) that, when annealed, gave rise to a two nucleotide 3' overhang (with the last annealed nucleotide being a G). After Cy5 labeling, the probe was purified using Illustra MicroSpin G-25 columns (GE Healthcare Life Sciences).

Full-length *BPC1, AZF1* cDNAs were cloned into pET32a and transformed into in *E. coli* (BL21). Protein was isolated as previous described[69]. For binding assays, a 20 µL reaction containing 2 µL of the protein extract, 3 µL of 1 pg/uL probe and 2 µL of 10× binding buffer (0.1 M Tris-HCI at pH 7.5, 0.5 M NaCl, 10 mM DTT, 10 mM EDTA, 50% glycerol, 0.5 mg/mL poly(dI-dC),1 mg/mL BSA) was used. For the AZF1 Zn-finger TF, the binding reaction was supplemented with 1mM CaCl2 and 0.1 mM ZnCl2. Free and bound probes were separated on a 6% PAGE gel in 0.5× TBE at 100 V for 60 min. The gel was scanned by a Typhoon scanner (Typhoon 9410 variable mode imager, Amersham) at BP 670 with a gain of 600. Shifted and unshifted probes were quantified using Image J. Full length EMSA gel images are provided in Supplementary Figure S13.

Tethering assay

BPC1 and AZF1 were cloned into LexA_DBD and Gal4_DBD vectors, respectively, as effectors[70]. The reporter contained 2 repeats each of the LexA and Gal4 binding sites with a 35S minimal promoter driving β-glucuronidase (GUS)[70]. 35S: LUC (firefly luciferase) was included to monitor transfection efficiency. Empty Gal4_DBD or Gal4_DBD_VP16 and LexA_DBD_VP16 served as controls. 16 hrs after transfection, $2x10^5$ protoplasts were used for ChIP (anti-FIE antiserum)[34].

Y1H screen

Functionally defined PREs (AG, SEP3, At5g) were cloned as 300bp DNA fragments into a modified pLacZi vector (Clontech) carrying *gLUC* instead of *LacZ* as reporter gene and integrated into the genome of yeast strain YM4271. A robotic Y1H screen was carried out against 1956 Arabidopsis TFs as previously described but using luciferase activity as readout[19]. The identified PRE-interacting TFs were further filtered on binding to multiple PRE fragments and binding strength.

For identification of telobox motif binding transcription factors, a telobox motif was inserted into a 30 bp region from a negative control fragment (NC_3; At3g18780 Actin2; Chr3: 6,476,550-6,476,579). Three copies of the 30 bp NC_3 region with telobox motif were inserted into bait vector pAbAi (Clontech), followed by integration into the genome of yeast strain Y1HGold (Clontech). Y1H screening against a TF library with about 1400 transcription factors[71] was carried out according to the Matchmaker® Gold Yeast One-Hybrid Library Screening System manual (Clontech), using Aureobasidin A resistance to select for binding.

Phylogenetic analyses

For phylogenetic trees, amino acid sequence alignment and generation of a neighbor-Joining (NJ) phylogenetic tree was performed using MEGA[72]with default settings. Phylogenetic shadowing was performed essentially as previously described[61], except that conserved regions were aligned using Clustal Omega (EMBL-EBI).

Statistical analysis:

Statistical tests performed, sample size and P-values are indicated in each figure legend. The investigator was not blinded to the group allocations during the experiment and variation was not estimated within each group of data. Dependent variables were continuous and all data points analyzed were independent. Throughout, the Kolmogorov–Smirnov[73] test was used to assess whether the data were normally distributed. For normally distributed data an unpaired one tailed *t*-test was used. In all other cases non-parametric tests were employed for two group comparisons (Mann-Whitney *U*-test[74]) and for multiple group comparison (Kruskal–Wallis test[75] combined with the Dunn's posthoc[76] test). Variances in some of the groups compared differ for biological reasons. Rejecting the null hypothesis based on these tests for this type of data implies that one group stochastically dominates a second group, that is to say if a value X is randomly chosen from one group, and a value Y is randomly chosen from another, then the probability of X>Y is greater than the probability of Y>X. The sample size was chosen based on prior studies that

showed significant effects using similar samples sizes, for example see[57]. P-values for all tests performed, as well as additional statistical parameters are listed in Supplementary Table 6.

Data availability

ChIPchip (CLF) and ChIPseq (all others) data were deposited in the Gene Expression Omnibus (GEO) under series numbers GSE7065 (CLF) and GSE7063 (input), GSE84483 (FIE, H3K27me3, AZF1 and BPC1) and GSE95562 (FIE in wild type and in *BPC+ZnF*^{KD}).

Code availability

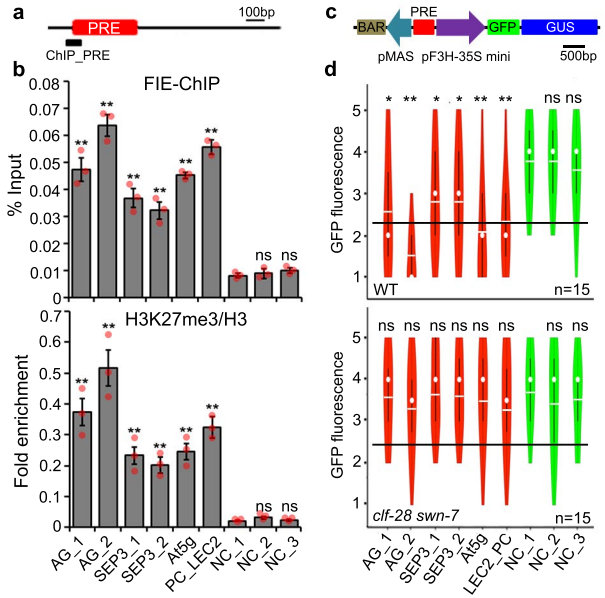
Motif prediction programs are available from their individual websites: Meme v4.9.1, Weeder v1.4.2, AlignAce v4.0, MotifSampler v3.2 and Bioprospector v2004. Motif enrichment and downstream analysis used the "Cistome"[77] pipeline, which is freely available as a web-based.

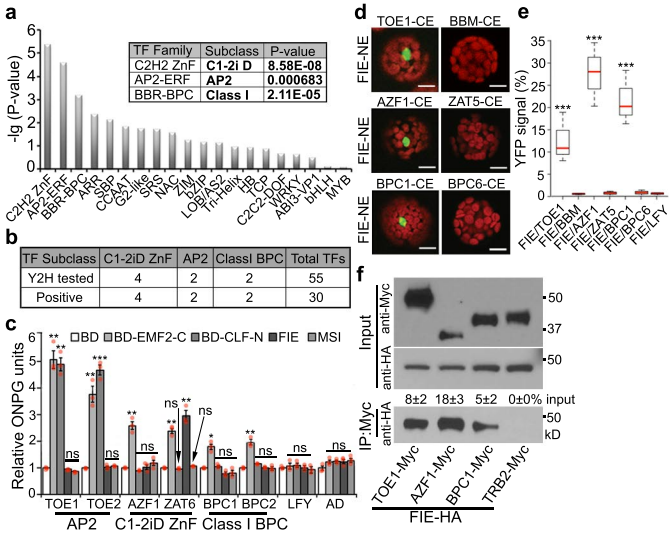
Methods only references

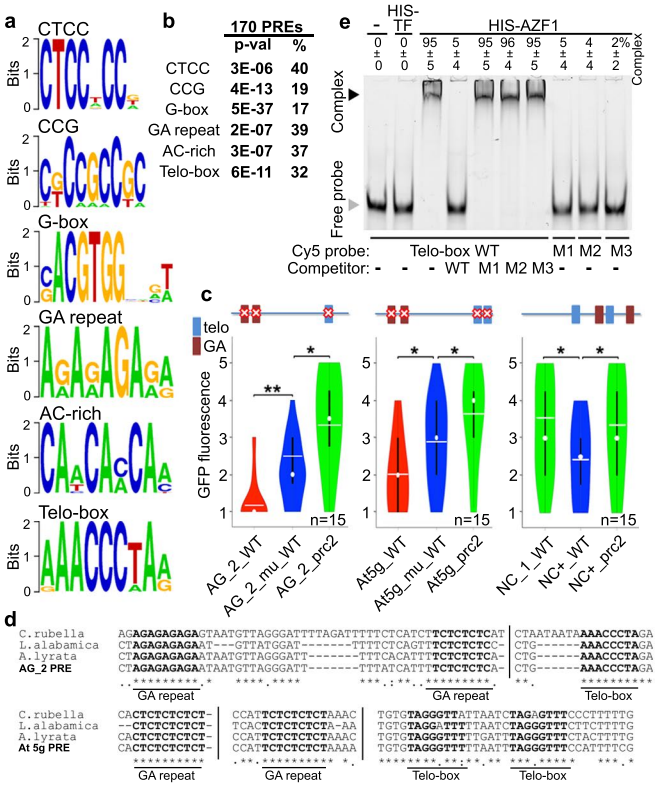
33. Lafos, M. *et al*. Dynamic regulation of H3K27 trimethylation during Arabidopsis differentiation. *PLoS genetics* **7**, e1002040 (2011).
34. Wood, C.-C. *et al*. The Arabidopsis thaliana vernalization response requires a polycomb-like protein complex that also includes VERNALIZATION INSENSITIVE 3. *Proc Natl Acad Sci U S A* **103**, 14631-14636 (2006).
35. Kim, S.-Y., Zhu, T. & Sung, Z.-R. Epigenetic regulation of gene programs by EMF1 and EMF2 in Arabidopsis. *Plant Physiol* **152**, 516-528 (2010).
36. Alexandre, C., Moller-Steinbach, Y., Schonrock, N., Gruissem, W. & Hennig, L. Arabidopsis MSI1 is required for negative regulation of the response to drought stress. *Mol Plant* **2**, 675-687 (2009).
37. Schubert, D. *et al*. Silencing by plant Polycomb-group genes requires dispersed trimethylation of histone H3 at lysine 27. *EMBO J* **25**, 4638-4649 (2006).
38. Maghuly, F. *et al*. Stress regulated expression of the GUS-marker gene (uidA) under the control of plant calmodulin and viral 35S promoters in a model fruit tree rootstock: Prunus incisa x serrula. *J Biotechnol* **135**, 105-16 (2008).
39. Blazquez, M. Quantitative GUS Activity Assay of Plant Extracts. *CSH Protoc* **2007**, pdb prot4690 (2007).
40. Mathieu, J., Yant, L.-J., Murdter, F., Kuttner, F. & Schmid, M. Repression of flowering by the miR172 target SMZ. *PLoS Biol* **7**, e1000148 (2009).
41. Murashige, T. & Skoog, F. A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiologia Plantarum* **15**, 473-497 (1962).
42. Clough, S.-J. & Bent, A.-F. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J* **16**, 735-743 (1998).
43. Nakagawa, T. *et al*. Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J Biosci Bioeng* **104**, 34-41 (2007).
44. Earley, K.-W. *et al*. Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J* **45**, 616-629 (2006).
45. Yamaguchi, A. *et al*. The microRNA-regulated SBP-Box transcription factor SPL3 is a direct upstream activator of LEAFY, FRUITFULL, and APETALA1. *Dev Cell* **17**, 268-278 (2009).

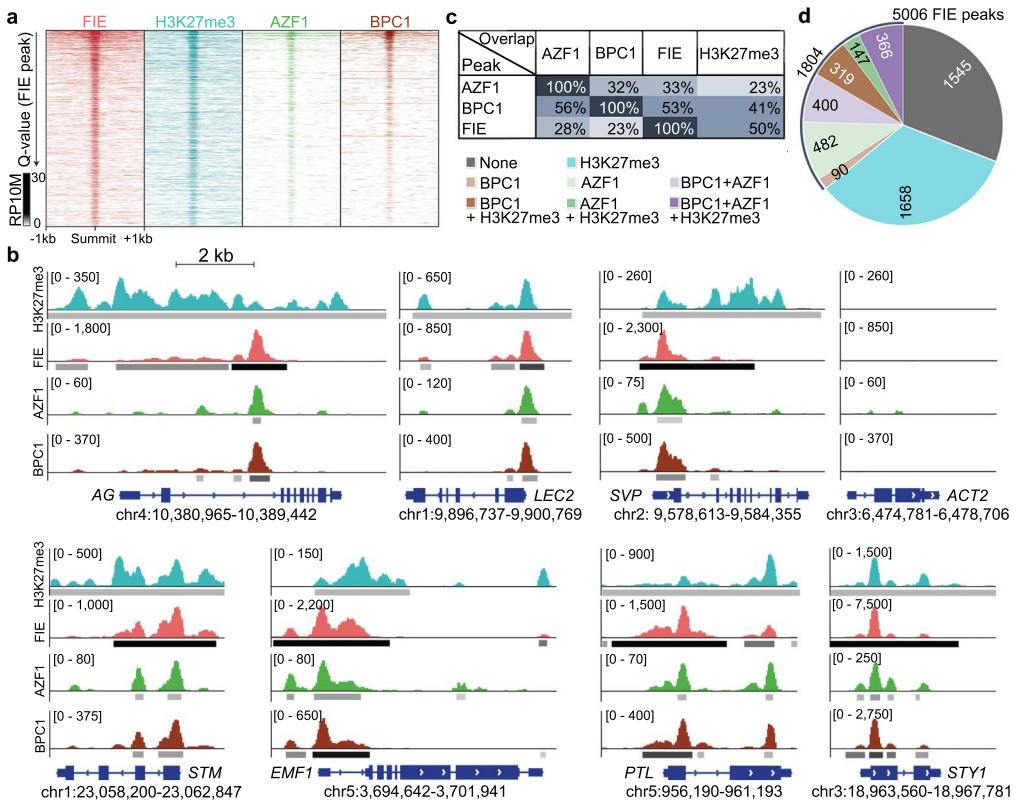
46. Zhang, X. *et al*. Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* **126**, 1189-1201 (2006).
47. Ji, H. & Wong, W.-H. TileMap: create chromosomal map of tiling array hybridizations. *Bioinformatics* **21**, 3629-3636 (2005).
48. Schmid, M. *et al*. A gene expression map of Arabidopsis thaliana development. *Nat Genet* **37**, 501-506 (2005).
49. Bailey, T.-L. & Elkan, C. The value of prior knowledge in discovering motifs with MEME. *Proc Int Conf Intell Syst Mol Biol* **3**, 21-29 (1995).
50. Hughes, J.-D., Estep, P.-W., Tavazoie, S. & Church, G.-M. Computational identification of cis-regulatory elements associated with groups of functionally related genes in Saccharomyces cerevisiae. *J Mol Biol* **296**, 1205-1214 (2000).
51. Pavesi, G., Mauri, G. & Pesole, G. An algorithm for finding signals of unknown length in DNA sequences. *Bioinformatics* **17 Suppl 1**, S207-14 (2001).
52. Hu, J., Li, B. & Kihara, D. Limitations and potentials of current motif discovery algorithms. *Nucleic Acids Res* **33**, 4899-913 (2005).
53. Quinlan, A.-R. & Hall, I.-M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
54. Chanvivattana, Y. *et al*. Interaction of Polycomb-group proteins controlling flowering in Arabidopsis. *Development* **131**, 5263-5276 (2004).
55. Wang, H. *et al*. Yeast two-hybrid system demonstrates that estrogen receptor dimerization is ligand-dependent in vivo. *J Biol Chem* **270**, 23322-23329 (1995).
56. Gehl, C., Waadt, R., Kudla, J., Mendel, R.-R. & Hansch, R. New GATEWAY vectors for high throughput analyses of protein-protein interactions by bimolecular fluorescence complementation. *Mol Plant* **2**, 1051-1058 (2009).
57. Wu, M.-F. *et al*. Auxin-regulated chromatin switch directs acquisition of flower primordium founder fate. *Elife* **4**, e09269 (2015).
58. Yoo, S.-D., Cho, Y.-H. & Sheen, J. Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat Protoc* **2**, 1565-1572 (2007).
59. Xiao, J. *et al*. O-GlcNAc-mediated interaction between VER2 and TaGRP2 elicits TaVRN1 mRNA accumulation during vernalization in winter wheat. *Nat Commun* **5**, 4572 (2014).
60. Yamaguchi, N. *et al*. PROTOCOLS: Chromatin Immunoprecipitation from Arabidopsis Tissues. *Arabidopsis Book* **12**, e0170 (2014).
61. Yamaguchi, N. *et al*. A molecular framework for auxin-mediated initiation of flower primordia. *Dev Cell* **24**, 271-282 (2013).
62. Aichinger, E. *et al*. CHD3 proteins and polycomb group proteins antagonistically determine cell identity in Arabidopsis. *PLoS Genet* **5**, e1000605 (2009).
63. Yamaguchi, N. *et al*. Gibberellin acts positively then negatively to control onset of flower formation in Arabidopsis. *Science* **344**, 638-41 (2014).
64. Li, K. *et al*. DELLA-mediated PIF degradation contributes to coordination of light and gibberellin signalling in Arabidopsis. *Nat Commun* **7**, 11868 (2016).
65. Langmead, B. & Salzberg, S.-L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
66. Zerbino, D.-R., Johnson, N., Juettemann, T., Wilder, S.-P. & Flicek, P. WiggleTools: Parallel processing of large collections of genome-wide datasets for visualization and statistical analysis. *Bioinformatics* **30**, 1008-1009 (2014).
67. Heinz, S. *et al*. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-89 (2010).
68. Du, Z., Zhou, X., Ling, Y., Zhang, Z. & Su, Z. agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res* **38**, W64-70 (2010).

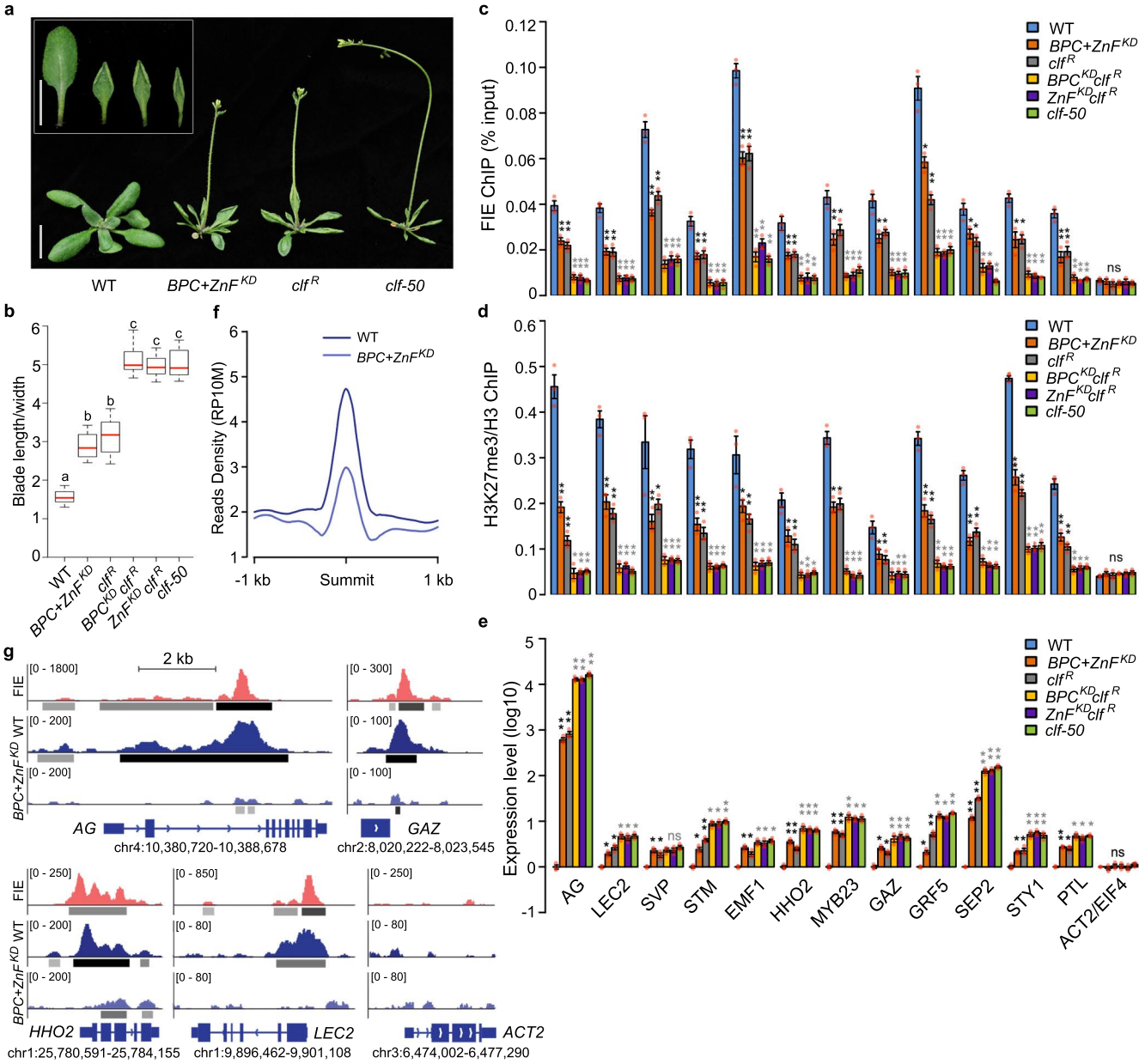
69. Zhang, B., Wang, L., Zeng, L., Zhang, C. & Ma, H. Arabidopsis TOE proteins convey a photoperiodic signal to antagonize CONSTANS and regulate flowering time. *Genes Dev* **29**, 975-987 (2015).
70. Tiwari, S.-B., Hagen, G. & Guilfoyle, T.-J. Aux/IAA proteins contain a potent transcriptional repression domain. *The Plant cell* **16**, 533-43 (2004).
71. Castrillo, G. *et al*. Speeding cis-trans regulation discovery by phylogenomic analyses coupled with screenings of an arrayed library of Arabidopsis transcription factors. *PLoS One* **6**, e21524 (2011).
72. Tamura, K. *et al*. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731-9 (2011).
73. Massey, F.J. The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association* **46**, 68–78 (1951).
74. Mann, H.-B. & Whitney, D.-R. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Annals of Mathematical Statistics* **18**, 50-60 (1947).
75. Kruskal, W.-H. & Wallis, W.-A. Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association* **47**, 583-621 (1952).
76. Dunn, O.J. Multiple comparisons using rank sums. *Technometrics* **6**, 241–252 (1964).
77. Austin, R.-S. *et al*. New BAR Tools for Mining Expression Data and Exploring Cis-Elements in Arabidopsis thaliana. *Plant J* (2016).

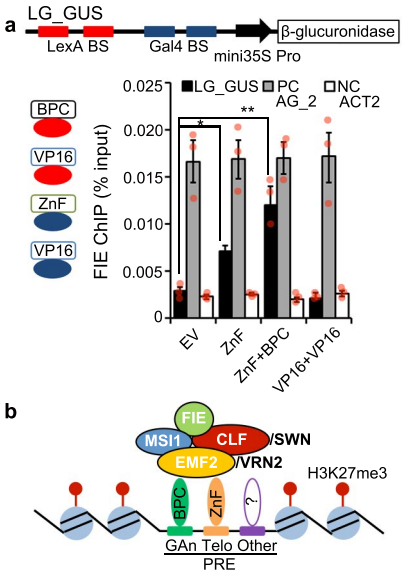

a
PRE
100bp
ChIP_PRE
b
FIE-ChIP
% Input
H3K27me3/H3
Fold enrichment
ns
AG_1
AG_2
SEP3_1
SEP3_2
At5g
PC_LEC2
NC_1
NC_2
NC_3
c
BAR
PRE
GFP
GUS
pMAS
pF3H-35S mini
500bp
d
GFP fluorescence
WT
n=15
clf-28 swn-7
LEC2_PC

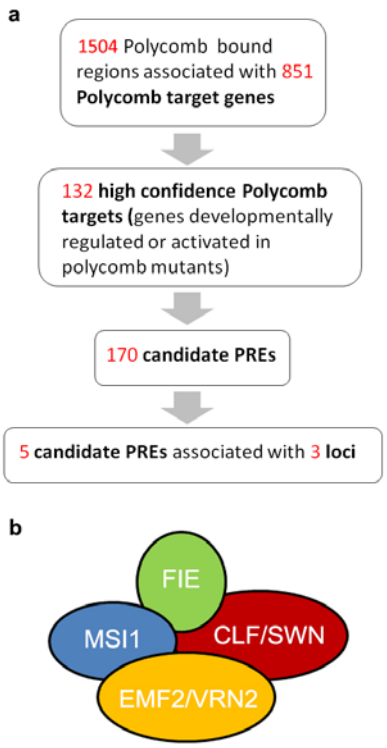

a
-lg (P-value)
TF Family | Subclass | P-value
C2H2 ZnF | C1-2i D | 8.58E-08
AP2-ERF | AP2 | 0.000683
BBR-BPC | Class I | 2.11E-05
C2H2 ZnF
AP2-ERF
BBR-BPC
ARR
SBP
CCAAT
G2-like
SRS
NAC
ZIM
ZF-HD
LOB-AS2
Tri-Helix
HB
TCP
C2C2-DOF
WRKY
ABI3-VP1
bHLH
MYB
b
TF Subclass | C1-2iD ZnF | AP2 | ClassI BPC | Total TFs
Y2H tested | 4 | 2 | 2 | 55
Positive | 4 | 2 | 2 | 30
c
BD BD-EMF2-C BD-CLF-N FIE MSI
Relative ONPG units
TOE1 TOE2 AZF1 ZAT6 BPC1 BPC2 LFY AD
AP2 C1-2iD ZnF Class I BPC
d
TOE1-CE
BBM-CE
AZF1-CE
ZAT5-CE
BPC1-CE
BPC6-CE
FIE-NE
e
YFP signal (%)
FIE/TOE1
FIE/BBM
FIE/AZF1
FIE/ZAT5
FIE/BPC1
FIE/BPC6
FIE/LFY
f
Input
anti-Myc
anti-HA
8±2 18±3 5±2 0±0% input
IP-Myc
kD
TOE1-Myc
AZF1-Myc
BPC1-Myc
TRB2-Myc
FIE-HA


a
CTCC
CCG
G-box
GA repeat
AC-rich
Telo-box
Bits
b
170 PREs
p-val %
CTCC 3E-06 40
CCG 4E-13 19
G-box 5E-37 17
GA repeat 2E-07 39
AC-rich 3E-07 37
Telo-box 6E-11 32
e
HIS-TF
HIS-AZF1
Complex
Free probe
Cy5 probe: Telo-box WT M1 M2 M3
Competitor: - - - WT M1 M2 M3 - - -
c
telo
GA
GFP fluorescence
n=15
AG_2_WT
AG_2_mu_WT
AG_2_prc2
At5g_WT
At5g_mu_WT
At5g_prc2
NC_1_WT
NC+_WT
NC+_prc2
d
C.rubella
L.alabamica
A.lyrata
AG_2 PRE
At 5g PRE
GA repeat
Telo-box


a
FIE
H3K27me3
AZF1
BPC1
Q-value (FIE peak)
RP10M
30
0
-1kb
Summit
+1kb
c
Overlap
Peak
AZF1
BPC1
FIE
H3K27me3
AZF1 100% 32% 33% 23%
BPC1 56% 100% 53% 41%
FIE 28% 23% 100% 50%
None
H3K27me3
BPC1
AZF1
BPC1+AZF1
BPC1 + H3K27me3
AZF1 + H3K27me3
BPC1+AZF1 +H3K27me3
d
5006 FIE peaks
1545
1658
90
482
400
1804
319
147
366
b
2 kb
H3K27me3
FIE
AZF1
BPC1
[0 - 350]
[0 - 1,800]
[0 - 60]
[0 - 370]
[0 - 650]
[0 - 850]
[0 - 120]
[0 - 400]
[0 - 260]
[0 - 2,300]
[0 - 75]
[0 - 500]
[0 - 260]
[0 - 850]
[0 - 60]
[0 - 370]
AG
LEC2
SVP
ACT2
chr4:10,380,965-10,389,442
chr1:9,896,737-9,900,769
chr2: 9,578,613-9,584,355
chr3:6,474,781-6,478,706
[0 - 500]
[0 - 1,000]
[0 - 80]
[0 - 375]
[0 - 150]
[0 - 2,200]
[0 - 80]
[0 - 650]
[0 - 900]
[0 - 1,500]
[0 - 70]
[0 - 400]
[0 - 1,500]
[0 - 7,500]
[0 - 250]
[0 - 2,750]
STM
EMF1
PTL
STY1
chr1:23,058,200-23,062,847
chr5:3,694,642-3,701,941
chr5:956,190-961,193
chr3:18,963,560-18,967,781


a
WT
BPC+ZnF KD
clf R
clf-50
b
Blade length/width
WT
BPC+ZnF KD
clf R
BPC KD clf R
ZnF KD clf R
clf-50
f
WT
BPC+ZnF KD
Reads Density (RP10M)
-1 kb
Summit
1 kb
c
FIE ChIP (% input)
WT
BPC+ZnF KD
clf R
BPC KD clf R
ZnF KD clf R
clf-50
ns
d
H3K27me3/H3 ChIP
e
Expression level (log10)
AG
LEC2
SVP
STM
EMF1
HHO2
MYB23
GAZ
GRF5
SEP2
STY1
PTL
ACT2/EIF4
g
2 kb
FIE
WT
BPC+ZnF KD
AG
chr4:10,380,720-10,388,678
GAZ
chr2:8,020,222-8,023,545
HHO2
chr1:25,780,591-25,784,155
LEC2
chr1:9,896,462-9,901,108
ACT2
chr3:6,474,002-6,477,290


a
LG_GUS
LexA BS
Gal4 BS
mini35S Pro
β-glucuronidase
BPC
VP16
ZnF
VP16
FIE ChIP (% input)
0.025
0.020
0.015
0.010
0.005
0
LG_GUS
PC
NC
AG_2
ACT2
*
**
EV
ZnF
ZnF+BPC
VP16+VP16
b
FIE
MSI1
CLF
/SWN
EMF2
/VRN2
H3K27me3
BPC
ZnF
?
GAn Telo Other
PRE

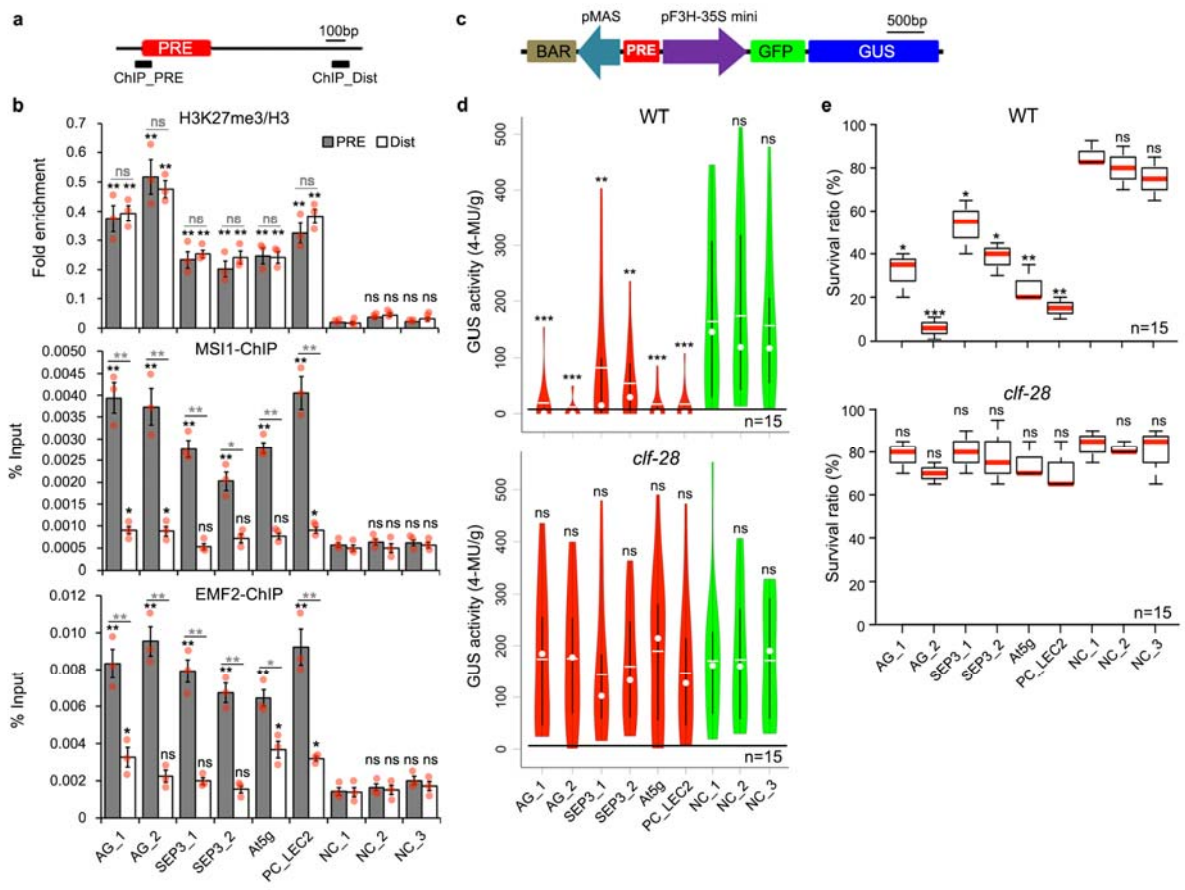


Supplementary Figure 1

High confidence PRC2 targets and candidate PREs.

(a) Flowchart for identification of candidate Arabidopsis PREs. We identified 1504 genomic regions marked by at least 3 of the following: H3K27me3, FIE, CLF or EMF1[15-17] (CLF ChIP-chip data: GSE7065) and linked these to 851 genes as previously described[23]. 132 of the 851 genes were significantly upregulated in *prc2* mutants[15,33] or strongly developmentally regulated[48] and thus considered high confidence PRC2 regulated genes. 170 candidate PREs were associated with the 132 genes. From the 170 PREs we selected 5 associated with 3 genes for test of PRE activity. See methods for additional details.

(b) Composition of the PRC2 complexes in the Arabidopsis sporophyte (diploid generation). Complex components: one of two SET domain methyltransferases (CURLY LEAF (CLF) or SWINGER (SWN)), one of two VEFS domain proteins (EMBRYONIC FLOWER2 (EMF2) or VERNALIZATION2 (VRN2)), a WD40 domain protein that can recognize H3K27me3 (FERTILIZATION INPENDPENT ENDOSPERM (FIE)) and a histone binding protein (MSI1)[3].



Supplementary Figure 2

Test of PRE activity.

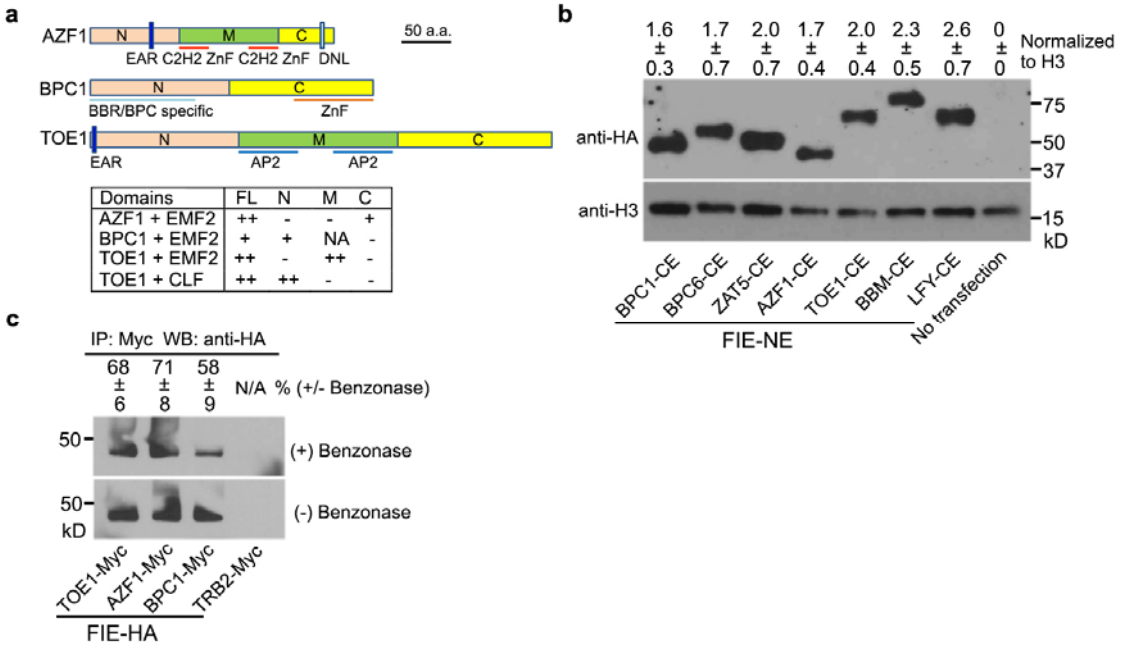
(a) Construct to test ability of candidate PREs or control DNA fragments (NC) to recruit PRC2 and H3K27me3. Below: regions tested: PRE or a distal site (Dist).

(b) H3K27me3/H3 abundance (top) and occupancy of PRC2 components MSI1 (middle) and EMF2 (bottom) assessed by ChIP-qPCR. Mean ± SEM from three experiments (red dots). Black asterisks- significantly different relative to NC_1; grey asterisks – significantly different occupancy at the PREs relative to the distal site (Dist). * P<0.05; ** P< 0.01, ns P>0.05; one-tailed unpaired student *t* test.

(c) Construct to test PRE-mediated reporter silencing.

(d) Fluorometric assay of beta-glucuronidase (GUS) activity of 15 independent transformants in the wild type (WT) (top) or a *PRC2* mutant (*clf-28*) (bottom). Violin plot of GUS activity in PREs (left) or negative controls (right): range, median= white circle, mean= white line, lower to higher quartile= vertical black line. Black bar: median GUS activity of all PRE populations in the wild-type background. * P<0.05; ** P<0.01; *** P<0.001; ns, not significant (P>0.25) relative to NC_1, one-tailed Mann-Whitney *U* test.

(e) Herbicide resistance (survival rate) conferred by the *BAR* gene product in n=60 independent T1 plants in the wild-type (top) or in the *prc2* (*clf-28*) mutant (bottom) background. Box and whisker plot with median (red line), upper and lower quartile (box) and minima, maxima (whiskers). P -value (one-tailed Mann–Whitney *U* test): ns, not significant; P> 0.07; * P<0.05; ** P<0.01; *** P<0.001 relative to NC_1.



Supplementary Figure 3

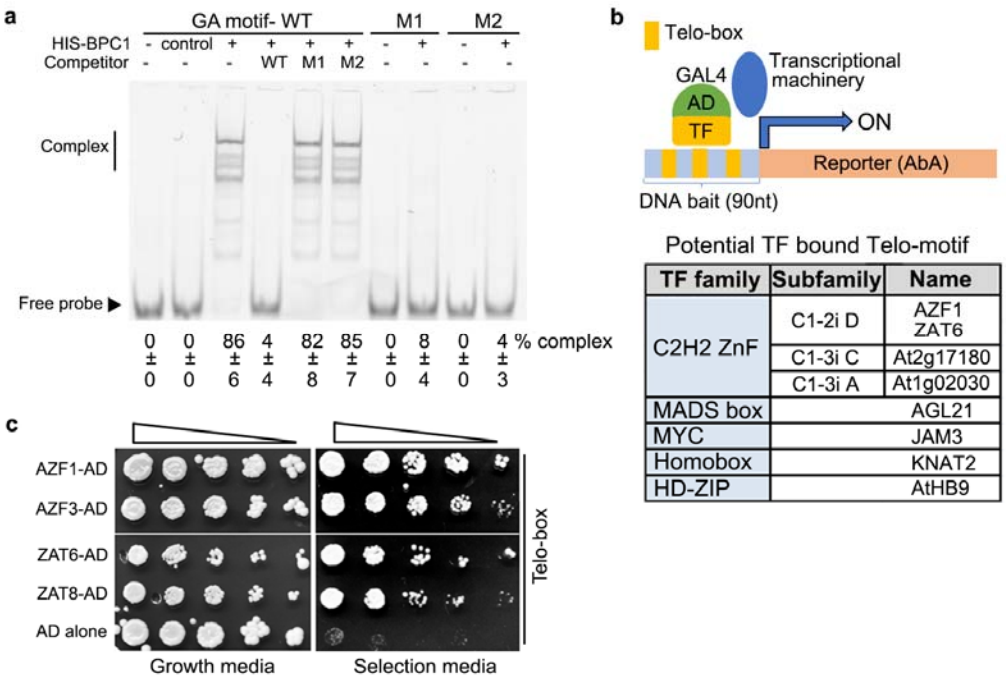
Physical interaction of PRE-binding transcription factors with PRC2.

(a) Domains of TFs important for interaction with PRC2. Yeast-two-hybrid interaction tests using EMF2 and CLF as bait and AZF1, BPC1 and TOE1 full-length proteins (FL), N-terminal domains (N), middle regions (M), or C-terminal domains (C) as prey. Known protein motifs[20,21,78] are shown below. NA = Not applicable. No single known protein motif is responsible for the TF/PRC2 interactions. a.a.: amino acid.

(b) Detection of TF abundance in plant cells co-transfected with FIE-NE and used for the BiFC analyses in Fig. 2d, e. Above: mean ± SEM of TF levels relative to those of histone H3 from three independent BiFC experiments.

(c) Co-IP of Myc-tagged TFs and PRC2 in the presence or absence of the benzonase endonuclease[79]. TOE1, AZF1 and BPC1 co-immunoprecipitate with PRC2 (HA-tagged FIE) in the absence and presence of the nuclease. Above: FIE co-IP in the presence relative to absence of benzonase, mean ± SEM from two independent co-IP experiments. The TRB2 sequence-specific binding protein serves as negative control. For expression levels of all 4 Myc-tagged proteins see Fig. 2f.

78. Wanke, D. *et al.* Alanine zipper-like coiled-coil domains are necessary for homotypic dimerization of plant GAGA-factors in the nucleus and nucleolus. *PLoS One* **6**, e16070 (2011).
79. Fiil, B.-K., Qiu, J.-L., Petersen, K., Petersen, M. & Mundy, J. Coimmunoprecipitation (co-IP) of Nuclear Proteins and Chromatin Immunoprecipitation (ChIP) from Arabidopsis. *CSH Protoc* **2008**, pdb prot5049 (2008).



Supplementary Figure 4

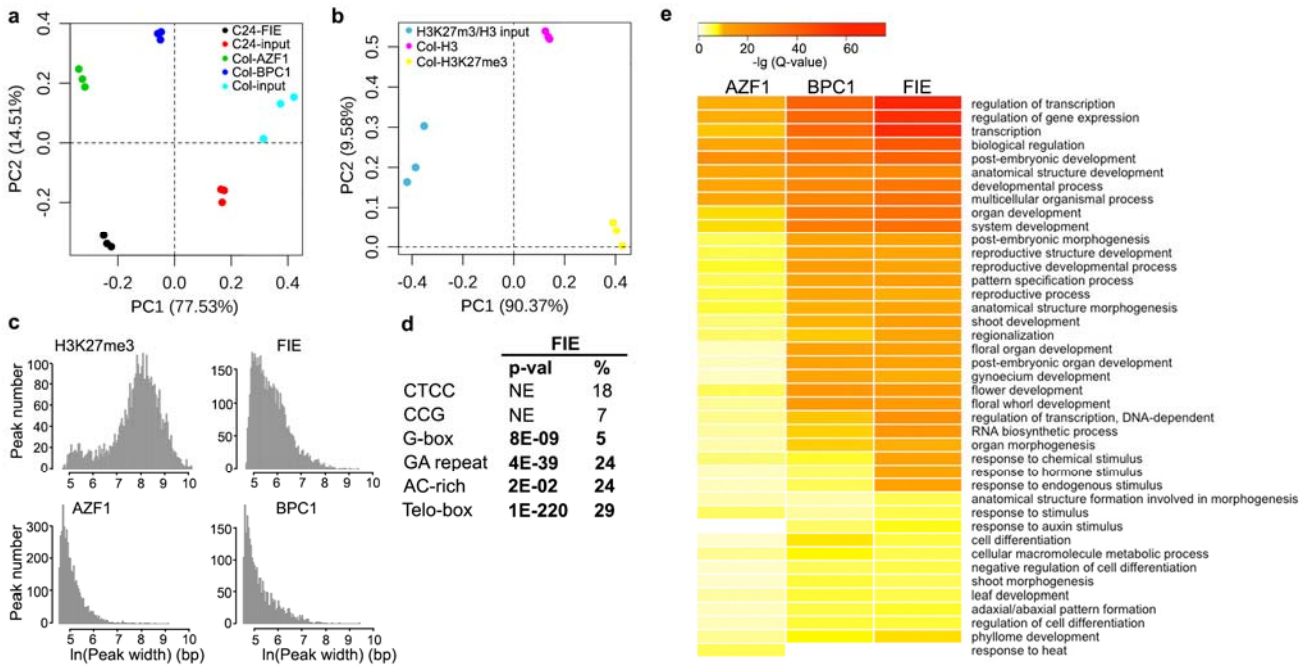
GA repeat and telobox motifs are bound by class I BPC and C1-2iD Zn Finger transcription factors.

(a) Electrophoretic mobility shift assay to test association of BPC1 with the GA repeats or with mutated versions thereof. WT, M1 and M2 were also used as cold competitors. Class I BPC TFs are known to oligomerize[22,28,78]. This, combined with the presence of two GA repeats in the tested DNA fragment, may explain why multiple shifted bands are observed when the protein is complexed with the DNA. Below: fraction of shifted DNA (% complex). Mean ± SEM of three EMSA experiments.

(b) Motif-based yeast one hybrid screen identified members of the C1-2iD family of Zn-finger TFs as telobox (**AAACCCTA**) binding transcription factors. The screen preferentially identified C2H2 ZnF proteins, in particular those of the C1-2iD subfamily (AZF1 and ZAT6).

(c) Confirmation of telobox interactome screen. Ten-fold serial dilutions of the yeast strains containing the bait DNA (b) integrated into the genome and plasmids containing the C1-2iD Zn-finger TFs indicated were plated on growth media (left) or on selection media (containing 600ng/ml aureobasidin A fungicide; right). See Supplementary Fig. 7 for a phylogenetic tree of ZnF TFs. The thin white line indicates where the plate image was cut.

78. Wanke, D. *et al.* Alanine zipper-like coiled-coil domains are necessary for homotypic dimerization of plant GAGA-factors in the nucleus and nucleolus. *PLoS One* **6**, e16070 (2011).



Supplementary Figure 5

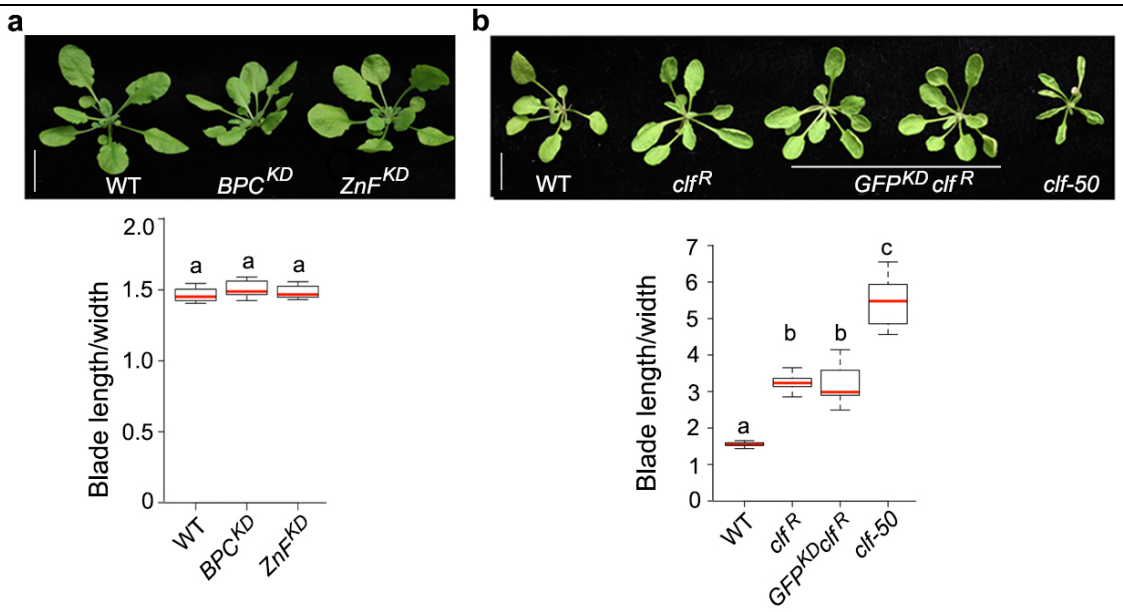
FIE, AZF1, BPC1 and H3K27me3 ChIPseq analysis.

(a, b) Principal component analysis (PCA) of RPM-normalized ChIP and input DNA reads for narrow (a) and broad (b) peak calling.

(c) Peak size distribution.

(d) Enrichment (p-values, converted from the Z-scores of the motif enrichment calculations using a normal distribution) and frequency (%) of PRE cis motifs under FIE-bound peaks. NE: not enriched (p>0.5).

(e) Functional classification of PRC2 (FIE), C1-2iD ZnF (AZF1) and class I BPC (BPC1) peak associated genes. Enrichment of Gene Ontology terms (FDR<10^{-5} in at least one of the datasets) for the genes associated with FIE, BPC1 and AZF1 peaks (Q<10^{-10}). The majority of the significant BPC1 and AZF1 target gene Gene Ontology terms are also significant Gene Ontology terms of FIE targets. Enriched GO terms include regulation of transcription, postembryonic development (reproductive development, shoot development, gynoecium development) and hormone response.

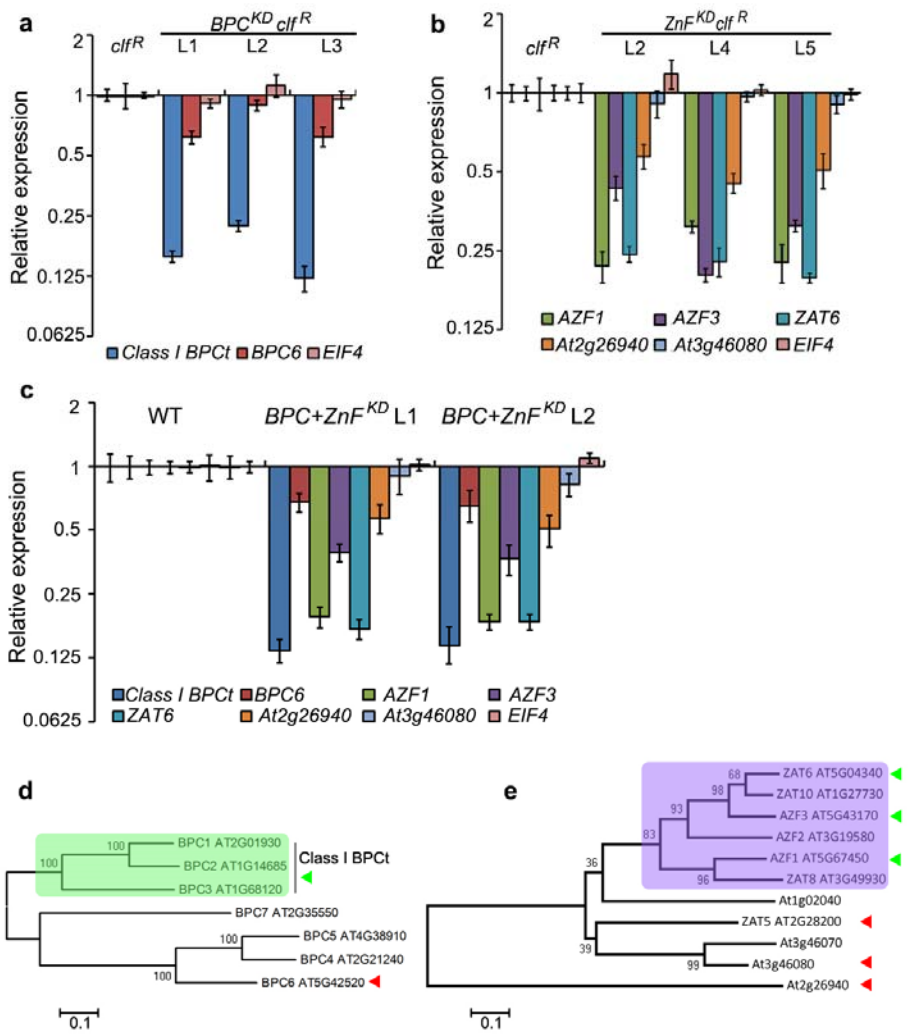


Supplementary Figure 6

Transcription factor family knockdown in the wild type and control knockdown in *clf*^{R}.

(**a**) class I BPC knockdown in the wild type (WT) does not cause leaf curling. Top: Representative images, bar = 1cm. Bottom: quantification of phenotypes. Box and whisker plot with median (red line, n=15 independent lines), upper and lower quartile (box) and minima, maxima (whiskers). Different letters above bars indicate significantly different groups, P<0.05 based on Kruskal-Wallis test with Dunn's posthoc test.

(**b**) Control knock-down (*GFP*^{KD}) in the hypomorph *clf*^{R} mutant does not enhance *clf*^{R} leaf curling. Top: Representative images, bar = 1cm. Bottom: quantification of phenotypes. Box and whisker plot with median (red line), upper and lower quartile (box) and minima, maxima (whiskers). Different letters above bars indicate significantly different groups, P<0.05 based on Kruskal-Wallis test with Dunn's posthoc test.



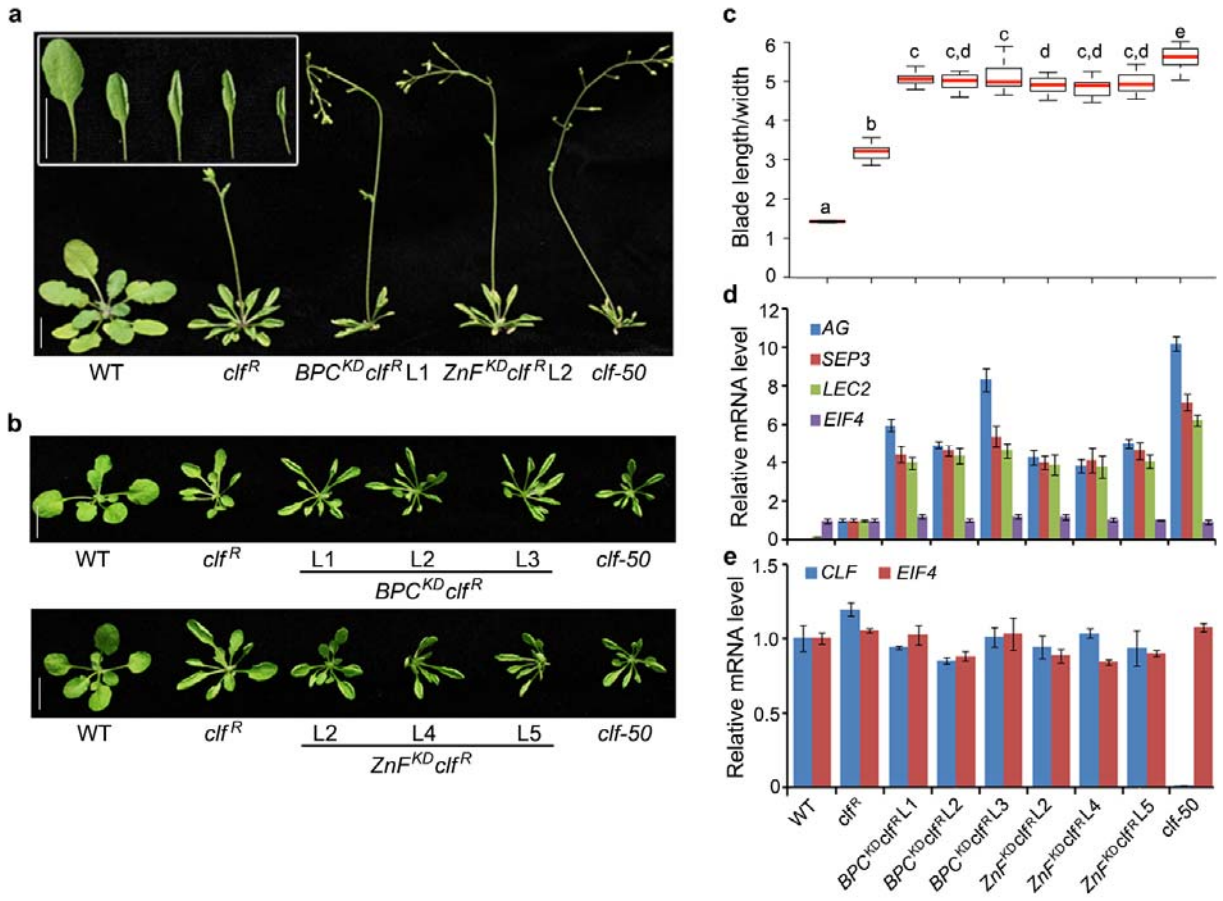
Supplementary Figure 7

RNAi-mediated knockdown of class I BPC and C1-2iD ZnF TF families.

(**a, b**) Class I BPC TF family knockdown by RNAi (*BPC*^{KD}; a) or C1-2iD C2H2 Zn-finger TF knockdown (*ZnF*^{KD}; b) was assayed in independent transgenic lines in the hypomorph *clf*^{R} mutant background. Expression of all genes tested is shown relative to that in *clf*^{R} (mean ± SEM) of one experiment for 3 independent lines. In each case, strong knockdown of the targeted TF family is observed (class I BPC and AZF1, AZF3, ZAT6), with more minor effects on distantly related BPC or ZnF TFs (BPC6, At2g26940, At3g46080). The translation initiation factor EIF4 served as qRT-PCR control.

(**c**) Simultaneous knockdown of Class I and C1-2iD C2H2 Zn TFs by RNAi (*BPC + ZnF*^{KD}) in the wild type tested as described in (a, b) in 2 independent lines.

(**d, e**) Phylogenetic tree of the BBR-BPC family of TFs (see also[22]) (d) and of select C2H2 Zinc finger TFs (see also[20]) plus a subset of additional C1-2i C2H2 ZnF proteins in Arabidopsis (e). Branch length is indicated. Light green shading highlights the class I BPC TFs and light purple shading the C1-2iD Zn-finger proteins targeted for knockdown. Triangles point to the genes tested by qRT-PCR in (a-c) or used as controls in BiFC (Fig. 2d,e; BPC6 and ZAT5). Triangle color indicates RNAi targets (green) or distantly related genes (red). Note that the primer pair employed for class I BPCt qRT-PCR amplifies all 3 genes in that clade.



Supplementary Figure 8

The effect of knockdown of class I BPC or C1-2iD Zn-finger TF family in the hypomorph *clf*^{R} mutant.

(**a**) Leaf curling (inset) and flowering time in wild type (WT), the weak *clf*^{R} mutant, class I BPC knockdown in *clf*^{R}, C1-2iD ZnF knock down in *clf*^{R} and in *clf-50* RNA null mutant, bar = 1cm.

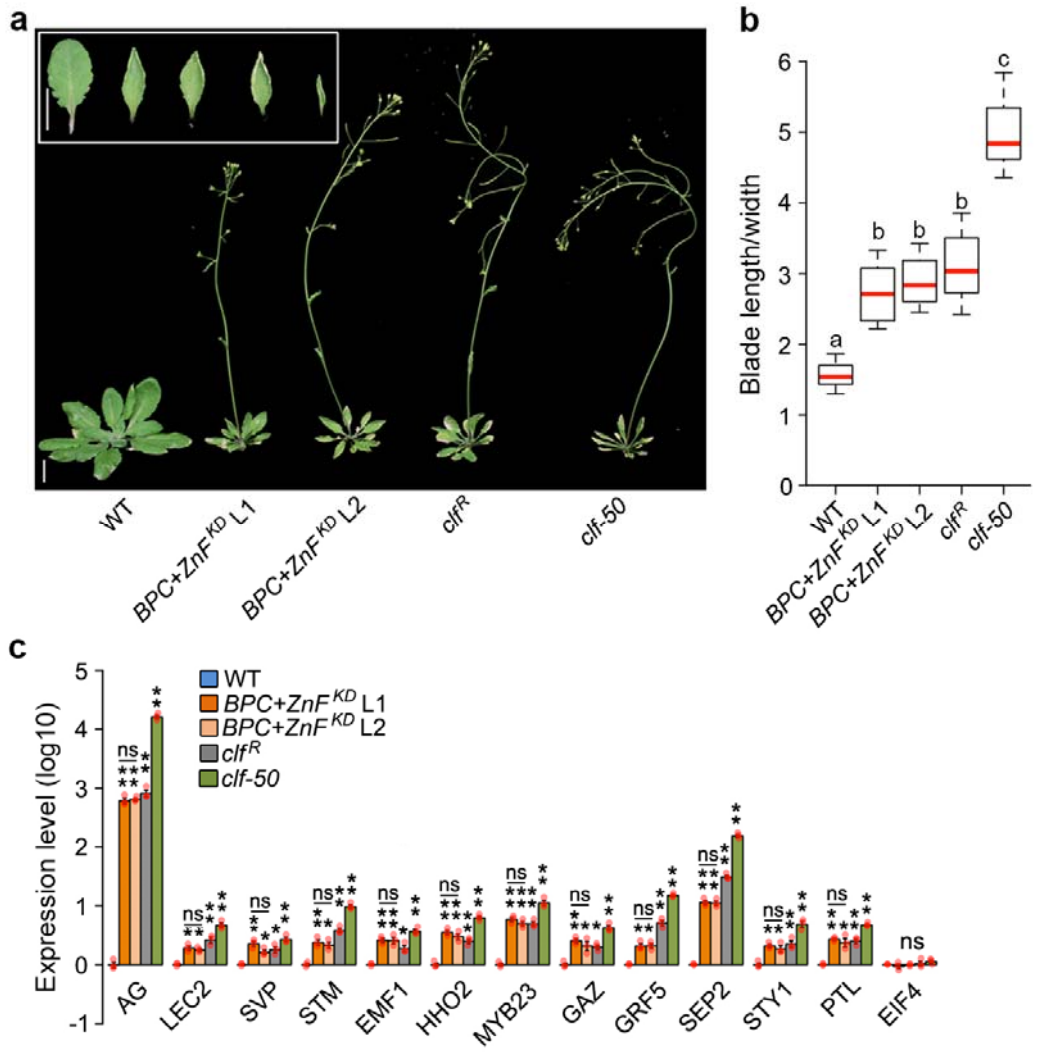
(b) Leaf curling phenotype of independent *BPC*^{KD} *clf*^{R} (top) and *ZnF*^{KD} *clf*^{R} (bottom) transgenic lines, bar = 1cm.

(c) Phenotype quantification of genotypes in (b). Box and whisker plot with median (red line, n=15 independent lines), upper and lower quartile (box) and minima, maxima (whiskers). Different letters above bars indicate significantly different groups ($p<0.05$) based on Kruskal-Wallis test with Dunn's posthoc test.

(d) De-repression of Polycomb target gene expression (*AG*, *SEP3*, *LEC2*) in the genotypes in (b). Expression relative to the parental line (*clf*^{R}) is shown. The housekeeping gene *EIF4* served as control.

(**e**) The abundance of the *CLF* mRNA relative to the wild type as measured by qRT-PCR. CLF levels are unchanged in the *BPC*^{KD}*clf*^{R} and the *ZnF*^{KD}*clf*^{R} plants shown in (b) compared to the parental line (WT). Shown are mean ± SEM of one experiment for three independent transgenic lines (d,e).

1



Supplementary Figure 9

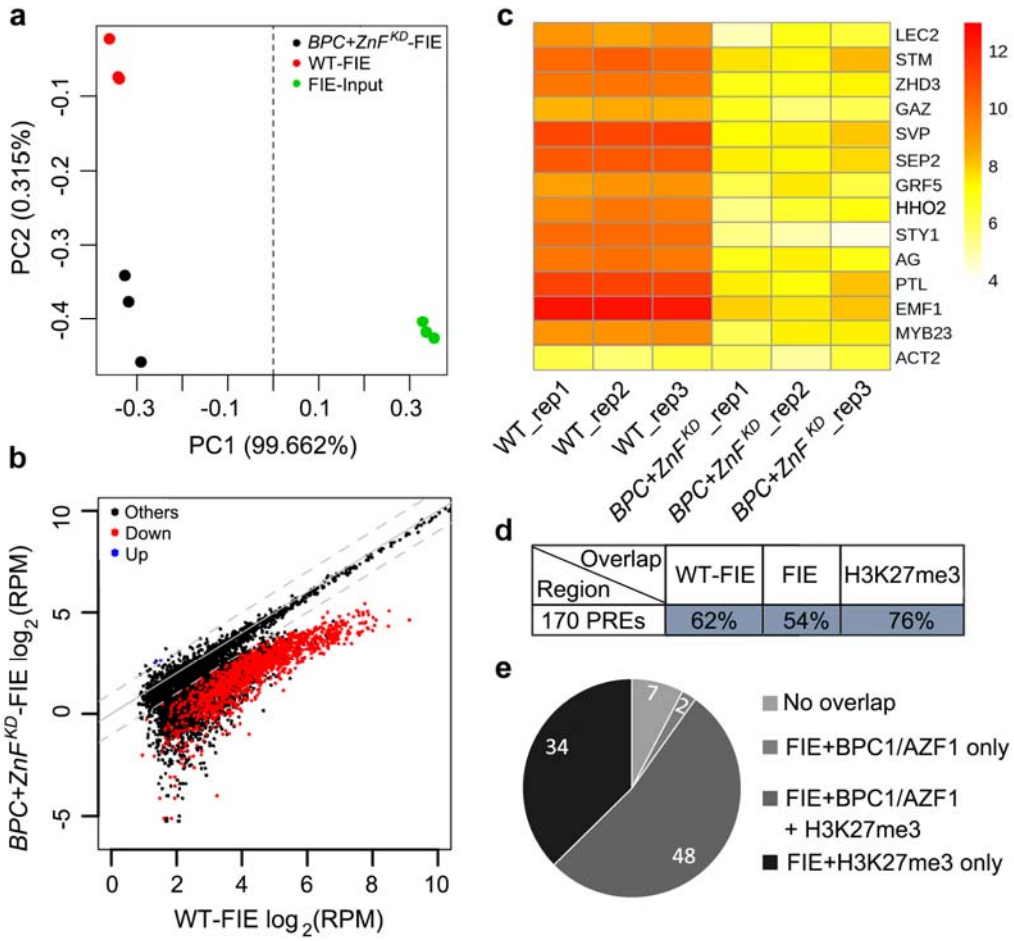
Phenotype of simultaneous knockdown of C1-2iD ZnF and class I BPC transcription factors.

(a) Leaf curling (inset) and early flowering in two independent *BPC*^{KD} *ZnF*^{KD} TF family knockdown lines, the hypomorph *clf*^{R} mutant and the null *clf-50* mutant relative to the wild type (WT), bar = 1cm.

(b) Quantification of the phenotypes in (a). Box and whisker plot with median (red line, n=15 independent lines), upper and lower quartile (box) and minima, maxima (whiskers). Different letters above bars indicate significantly different groups ($p<0.05$) based on Kruskal-Wallis test with Dunn's posthoc test.

(c) Misexpression of polycomb target genes in the genotypes listed in (a). Expression in the mutant lines is show relative to that of the wild type. Mean ± SEM from three experiments (red dots). * $P<0.05$; ** $P< 0.01$, ns $P>0.05$; relative to the wild type, one-tailed unpaired *t* test. The two independent *BPC*^{KD} *ZnF*^{KD} TF family knockdown lines are not significant different from each other, ns $P>0.05$; one-tailed unpaired *t* test.

2
3



Supplementary Figure 10

FIE ChIP in wild type or *BPC+ZnF KD* plants.

(**a**) Principal component analysis (PCA) of RPM-normalized ChIP and input DNA.

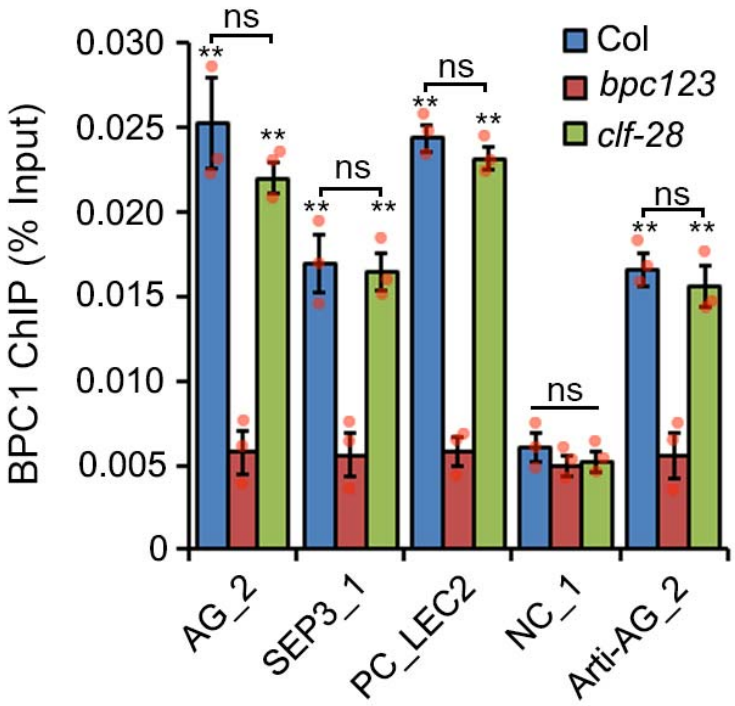
(b) Comparison of FIE ChIPseq reads from wild-type (WT) and *BPC+ZnF KD* plants. Shown are ChIPseq reads mapping to the WT-FIE peak regions. Regions with significantly (P-value <0.01 and 2-fold change) increased and decreased read density in *BPC+ ZnF KD* are indicated by blue and red color, respectively.

(c) Heatmap of RPM normalized read counts of FIE ChIPseq at the targets tested by qRT-PCR in Figure 5.

(d) Percent overlap table (by row) of 170 candidate PREs with significant ($Q < 10^{-10}$) ChIPseq peaks for FIE in the wild type background (from Figure 5), and FIE and H3K27me3 (from Figure 4). Shading highlights strength of overlap.

(**e**) Piechart for the overlap between significant ($Q<10^{-10}$) FIE, TF (BPC1/AZF1) and H3K27me3 peaks identified by ChIPseq in 30-hr-old plants and the 170 computationally defined candidate PREs from the vegetative phase of development.

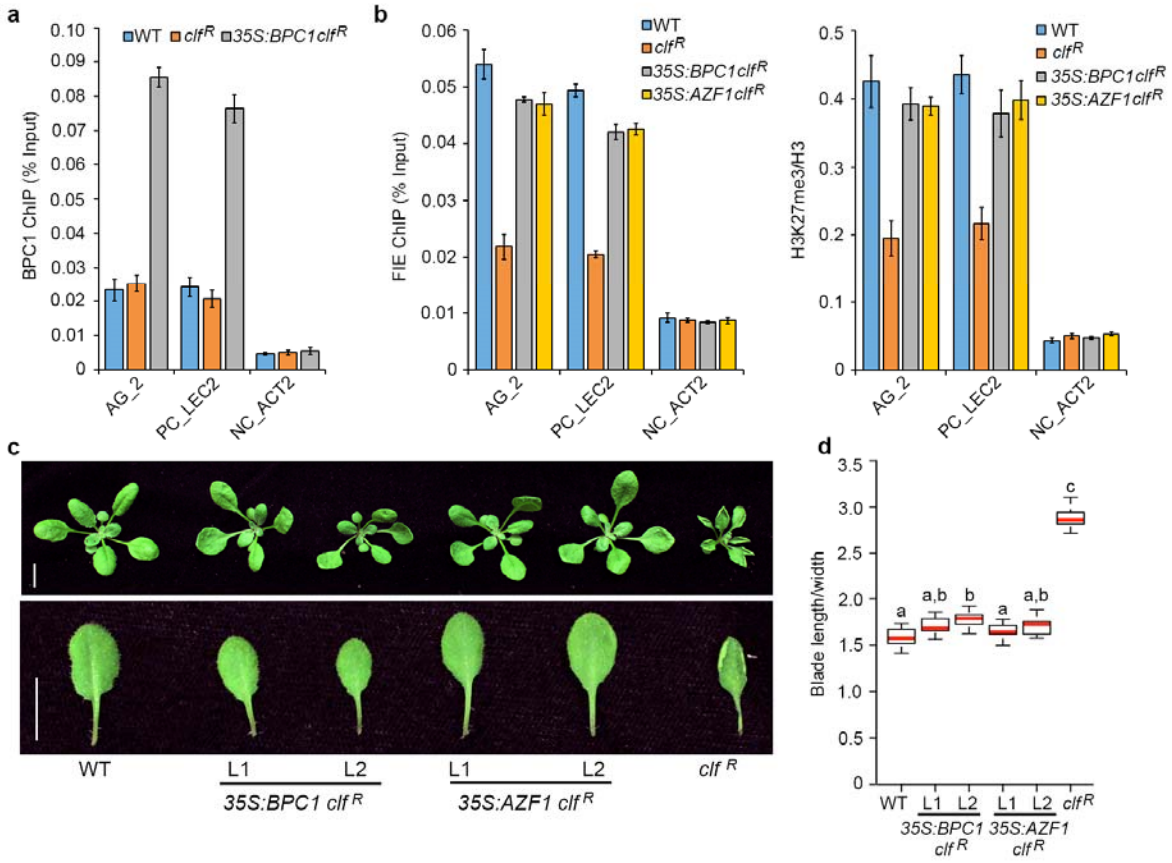
4



Supplementary Figure 11

BPC1 TF chromatin occupancy in *prc2* mutant.

ChIP using antiserum against class I BPC proteins to test BPC1 occupancy in the wild type (WT), the *bpc123* triple mutant and the *clf-28* (*prc2*) mutant. Binding was assayed at endogenous PREs (AG_2, SEP3_1, PC_LEC2), a control locus (NC_1) or a PRE reporter (Arti-AG_2). Shown are mean ± SEM of three experiments (red dots). ** P< 0.01, relative to *bpc123* mutant, one-tailed unpaired *t* test. ns, no significant difference (P>0.08) between BPC1 binding in Col and *clf-28*, one-tailed unpaired *t* test.



Supplementary Figure 12

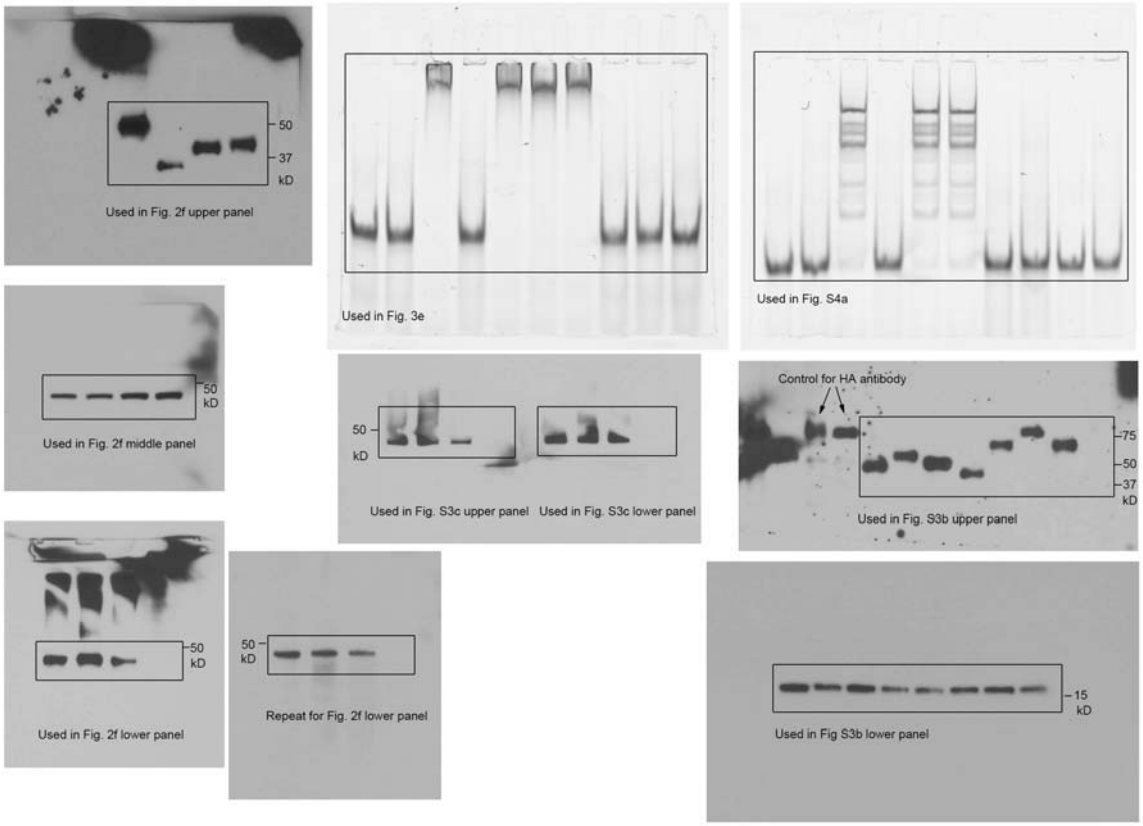
Gain-of BPC1 or AZF1 function rescues a hypomorph *prc2* mutant.

(a) ChIP using antiserum against class I BPC proteins (see Supplementary Fig. 11) to test BPC1 occupancy in the wild type (WT), the hypomorph *clf*^{R} (*prc2*) mutant and 35S:BPC1 *clf*^{R}. Binding was assayed at endogenous PREs (AG_2, PC_LEC2) or an endogenous control region (NC_ ACT2). Shown is the mean ± SEM of two independent TF overexpression lines.

(b) ChIP using antiserum against FIE (PRC2; left) or H3K27me3/H3 (right) in the wild type (WT), *clf*^{R}, 35S:BPC1 *clf*^{R} and 35S:AZF1 *clf*^{R}. Shown is the mean ± SEM of two independent TF overexpression lines.

(c, d) Leaf curling in wild type (WT), two independent lines (L1 and L2) of 35S:BPC1 *clf*^{R} and of 35S:AZF1 *clf*^{R} and the hypomorph *clf*^{R} mutant. (c) Representative images. Bar = 1cm. (d) Quantification of leaf curling in the genotypes shown in (c). Box and whisker plot with median (red line, n=15 independent lines), upper and lower quartile (box) and minima, maxima (whiskers). Different letters above bars indicate significantly different groups, $p<0.05$ based on Kruskal-Wallis test with Dunn's posthoc test.

7



Supplementary Figure 13

Full length gel images for all figures.

Left: Images for Co-IP. Two images are provided for the lower panel, the left image has non-specific signal at the top.
Center (top): Image for telobox EMSA; (below): Co-IP in the absence and presence of Benzonase.
Right (top): Image for GA repeat EMSA; (below): Quantification of proteins in protoplasts used for BiFC.